NBU Series in Cognitive Science

Advances in Analogy Research:

Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences

19980923 005

Edited by Keith Holyoak, Dedre Gentner, and Boicho Kokinov

AQ F98-12-2512

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 21 July 1998	3. REPORT TYPE AND DATES COVERED Conference Proceedings

4. TITLE AND SUBTITLE

International Workshop on Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences

5. FUNDING NUMBERS

F61775-98-WE105

6. AUTHOR(S)

Holyoak, Kieth; Dedre Gentner; and Boicho Kokinov, ed.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

New Bulgarian University
21 Montevideo St.
Sofia 1635
Bulgaria

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

EOARD
PSC 802 BOX 14
FPO 09499-0200

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

CSP 98-1069

11. SUPPLEMENTARY NOTES

ISBN 954-535-200-0

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

A

13. ABSTRACT (Maximum 200 words)

The Final Proceedings for International Workshop on Advances in Analogy Research, 17 July 1998 - 20 July 1998

Topics include Artificial Intelligence/Computational Modeling, Cognitive Psychology, Developmental Psychology, Neuropsychology, Philosophy, Cognitive Linguistics, as well as various applications in Education, Legal and Political Reasoning.

14. SUBJECT TERMS

DTIC QUALITY INSPECTED 1

Human Factors, psychology, artificial intelligence, cognition

15. NUMBER OF PAGES

419

16. PRICE CODE

N/A

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19, SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

NBU Series in Cognitive Science

Advances in Analogy Research:
Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences

Edited by Keith Holyoak, Dedre Gentner, and Boicho Kokinov

New Bulgarian University, Sofia, 1998

Copyright © 1998 by New Bulgarian University

All right reserved. No part of this publication may be reproduced , stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without permission of the publisher.

Printed in Bulgaria
ISBN: 954-535-200-0

Table of Contents

Keynote papers

Analogy in a Physical Symbol System - Keith Holyoak, John Hummel 9
The Role of Similarity in How we Categorize the World - James Hampton 19
Why Conceptual Combination is Seldom Analogy - Mark Keane, Fintan Costello 31
Analogical Problem-solving by Chimpanzees - David Oden, Roger Thompson, David Premack 38
Analogical Reasoning in Children - Usha Goswami 49
Relational Processing in Higher Cognition - Graeme Halford, William Wilson, Steven Phillips 57
Emotional Analogies and Analogical Inference - Paul Thagard 74
Analogy as a Cognitive Vehicle in Penetrating a New Domain - Adam Biela 86
Exploring Analogy in the Large - Kenneth Forbus 92
Comparison and Cognition - Dedre Gentner 94
Analogy is like Cognition: Dynamic, Emergent, and Context-Sensitive - Boicho Kokinov 96
Analogy as the Core of Cognition - Douglas Hofstadter *
Analogy in Problem Solving, from the Routine to the Creative - Jaime Carbonell*
Analogy and Conceptual Integration - Gilles Fauconnier*

Papers

Telling juxtapositions: Using Repetition and Alignable Difference in Diagram Understanding - Ronald Ferguson, Kenneth Forbus 109
Making Sense of Analogies in Metacat - James Marshall, Douglas Hofstadter 118
Mapping and Access in Analogy-Making: Independent or Interactive?
A Simulation Experiment with AMBR - Alexander Petrov , Boicho Kokinov 124
Principle Differences in Structure-Mapping: The Contribution of Non-Classical Models - Tony Veale, Mark Keane 135
An Argument for Derivational Analogy - Erica Melis, Jaime Carbonell 144
Structured Operations with Distributed Vector Representations - Tony Plate 154
Dual Role of Analogy in the Design of a Cognitive Computer - Pentti Kanerva 164
NetAB: A Neural Network Model of Analogy by Discovery - J. E. McCredden 171
Connections, Binding, Unification and Analogical Promiscuity-Ross Gayler, Roger Wales 181
Structure and Pragmatics in Analogical Inference - Arthur Markman, Adalis Sanchez 191
Pragmatic Effects on Speed of Analogical Problem Solving - Bruce Burns 201
Mapping Conceptual and Spatial Schemas - Merideth Gattis 210
Analogical Distance and Purpose in Creative Thought:
Mental Leaps versus Mental Hops - Thomas B. Ward 221
Reasoning by Analogy and Memory for Cases in the Game of Chess - Evelyne Cauzinille-Marméche, André Didierjean 231
Analogical Construction of a System of Simulation : a case study in physics - Sandra Bruno 237
The Copycat Project: Towards A Conceptual Fluidity Theory - Bruno Vivicorsi 245
Reasoning by Analogy: Representation of Pragmatic Information from Target Knowedge Influences Mapping - David Cayol 253

The Neuroanatomy of Analogical Reasoning - Charles Wharton, Jordan Grafman, Stephen Flitman, Eric Hansen, Jason Brauner, Allison Marks, Manabu Honda 260
Why Monkeys and Pigeons, Unlike Certain Apes, Cannot Reason Analogically - Roger Thompson, David Oden 269
The Effect of Language on Similarity: The Use of Relational Labels Improves Young Children's Performance in a Mapping Task - Mary Jo Rattermann, Dedre Gentner 274
Analogical Problem-Solving in Preschool Children - M. Bastien-Toniazzo, A. Blaye, D. Cayol 283
Concerning the Role of Analogy in Metaphor Processing - John Barnden 291
Alignment and Abstraction in Metaphor - Brian Bowdle 300
Evidence for Metaphoric Representation: Understanding Time - Lera Boroditsky 308
Metaphor: Shared Experience Structure or Cross-Domain Analogy? - Maciej Haman 320
Analogy Underlies Sentence Generation and Interpretation - Nili Mandelblit 328
Analogies in Design Activities: A Study of the Evocation of Intra- and Interdomain Sources - Nathalie Bonnardel 336
'Don't think, but look!': A Gestalt Interactionist Approach to Legal Thinking - Dan Hunter, Bipin Indurkhya 345
Depictive Analogies. - Barbara Tversky 354
Analogy and Induction : Which (missing) link ? - Antoine Cornuéjols, Jacques Ales-Bianchetti 365
Justification of Analogy by Abstraction - Hiroaki Suzuki 373
Analogy Making as a Categorization and an Abstraction Process - Emmanuel Sander, Jean-François Richard 381
Wittgenstein and the Ontological Status of Analogy - Michael Ramscar, Ulrike Hahn.......... 390

Abstracts

Multiplicative Binding, Representation Operators, and Analogy - Ross Gayler 404
Adaptation of Non-Isomorphic Sources in Analogical Problem Solving - Ute Schmid 406
"AQuARIUM: A Hierarchically-Supported Mono-Symbolic Language for Analogic Integration" - Ron Cottam, Willy Ranson, Roger Vounckx 408
The Mechanisms of Mapping: Evidence from On-line Analogy Judgments- Kenneth Kurtz, Dedre Gentner..410
An Experimental Paradigm to Study Spontaneous Analogies Involved in Problem Solving Situations - Emmanuel Sander, Evelyne Clement 411
Learning from Examples: Case-Based Reasoning in Chess for Novices - Andre Didierjean, Evelyne Cauzinille-Marmèche 413
Development of Numerical Equivalence Judgments: Appearances Count - Kelly Mix 414
The effects of a Training Program on the Analogical Reasoning Abilities of an Elementary School-Aged Sample - Doris Johnson 416
Analogical Generalism: An Analogical Perspective on the Evolution of Language - Jamie Carnie..417
Metaphors in Mind and Discourse - Elena Andonova 419

Preface

This volume is a result from the collective efforts of many researchers who participated in an interdisciplinary workshop on "Advances in Analogy Research" held in July 1998 at the Central and Eastern European Center for Cognitive Science at the New Bulgarian University, Sofia.

The purpose of the workshop has been to stimulate researchers in the field of analogy to cooperate more intensively and to integrate various approaches and data in their studies. Its aim has been to advance our understanding of the cognitive mechanisms of analogy-making, i.e. how people notice/perceive analogies, how they retrieve analogs from memory or how they construct them, how they map and transfer knowledge from one domain to another, how they combine knowledge from multiple analogs or how they combine analogy with rule-based reasoning, how they generalize and learn from the analogies made, how they use analogies for problem solving, explanation, argumentation, creation. What is the place of analogy among the various cognitive processes, such as perception, thinking, memory, learning, etc. What is the role of analogy in human development? Which are the brain structures involved in analogy-making processes? What kind of analogy-related deficits do brain-damaged patients exhibit?

This workshop has been highly interdisciplinary and has made a serious attempt to integrate the knowledge researchers have accumulated on analogy-making in various areas: Artificial Intelligence/Computational Modeling, Cognitive Psychology, Developmental Psychology, Neuropsychology, Philosophy, Cognitive Linguistics, as well as various applications in Design, Legal and Political Reasoning, Education, etc. A serious attempt has been made to integrate all the positive results obtained so far in theories of analogy-making, computational modeling, and experimental work.

This has been a unique workshop which drew together most of the key researchers in the field of analogy and gave them the chance to exchange ideas, share visions, and form friendships. The workshop has attracted about 70 participants from all over the world (25 participants from USA, 10 from France, 6 from Germany, 5 from UK, 3 from Australia, 3 from Ireland, 2 from Canada, 2 from Japan, 2 from Poland, 2 from Belgium, 1 from the Netherlands, 1 from Sweden, 1 from New Zealand, and 7 from Bulgaria). They presented 59 papers, including 14 key talks, 30 talks, and 15 posters.

We would like to thank especially all the key speakers and presenters for their valuable contributions to the success of the workshop. We would like also to thank the local organisers Guergana Yancheva, Iliana Haralanova, Ivailo Milenkov, Ivailo Panov.

We wish to thank the following for their contribution to the success of this workshop:
Cognitive Science Society - USA, Fulbright Commission - Sofia, MIT Press - USA, United States Air Force European Office of Aerospace Research and Development.

Dedre Gentner, Keith Holyoak, Boicho Kokinov

Keynote Papers

ANALOGY IN A PHYSICAL SYMBOL SYSTEM

Keith J. Holyoak [1,2], John E. Hummel [1]
Department of Psychology [1] and Brain Research Institute [2]
University of California, Los Angeles
Los Angeles, CA 90095-1563 USA
email: holyoak@lifesci.ucla.edu, jhummel@lifesci.ucla.edu

Abstract: Analogy, and relational reasoning in general, depend on a Phsyical Symbol System (PSS). We argue that the biological PSS that underlies human (an other primate) intelligence is based on mechanisms for dynamically and independently binding fillers to roles, which require working-memory representations maintained by dorsolateral prefrontal cortex. Our approach, termed symbolic connectionism, realizes symbolic processing in a neural network. The approach is instantiated in the LISA model (Hummel & Holyoak, 1997), which performs analog retrieval, mapping, inference, and schema induction. LISA makes a strong distinction between the driver analog, which is activated sequentially in small groups of propositions, and the recipient analog, which passively responds to the activity of the driver. The driver/recipient distinction leads to predictions about asymmetries and grouping effects in mapping, which we have tested and confirmed. More generally, the model is consistent with recent evidence that working-memory resources are required for more complex relational mappings, and that relational processing depends on the dorsolateral prefrontal cortex.

PHYSICAL SYMBOL SYSTEMS

A foundational principle of modern cognitive science is the Physical Symbol System hypothesis, which states simply that human cognition is the product of a physical symbol system (PSS). A symbol is a pattern that denotes something else; a symbol system is a set of symbols that can be composed into more complex structures by a set of relations. The term "physical" conveys that a symbol system can and must be realized in some physical way in order to create intelligence. The physical basis may be the circuits of an electronic computer, the neural substrate of a thinking biological organism, or in principle anything else that could implement a Turing machine-like computing device (Newell, 1980, 1990; Vera & Simon, 1993, 1994).

Because analogical thinking, like other forms of relational reasoning, depends on composed symbols (propositions specifying relations between the elements that fill specfic roles, where the elements may themselves be propositions), it necessarily requires a PSS. But what sort of cognitive architecture could implement a PSS? The fact that the mind performs symbol manipulation is important in constraining Marr's (1980) computational level, but it remains to be determined how the mind performs symbolic computation, which is a question at the level of representation and algorithm; and also how the PSS is realized in the brain, which is a question at the level of implementation. That is, the PSS that we seek to understand is that which is the product of biological evolution.

Both analogy and the PSS that underlies it appear to be late evolutionary developments. Relational processing appears to be a key innovation in primate intelligence (see Tomasello & Call, 1997); simple relational analogies can be solved by symbol-trained chimpanzees (Gillan, Premack & Woodruff, 1981; Premack, 1983), and more complex analogical reasoning is a uniquely human capability (Holyoak & Thagard, 1995). A great deal of evidence indicates that the prefrontal cortex is a key component of the neural substrate for the PSS (for reviews see

Grafman, Holyoak & Boller, 1995; Shallice & Burgess, 1991). Reasoning abilities and prefrontal cortex have developed in tandem across both phylogeny and ontogeny (Benson, 1993). Neuropsychological studies of frontal lobe function indicate that prefrontal cortical, especially in the dorsolateral prefrontal cortex (DLPFC), dysfunction leads to selective decrements in performance on a variety of complex cognitive tasks that depend on relational processing. The DLPFC is critical to working memory, to which relational reasoning appears to be intimately connected. In particular, an essential role of working memory in reasoning may be to maintain bindings between roles and fillers in relational representations (Robin & Holyoak, 1995). Thus, the DLPFC may be a major component of the neural system that implements the PSS, and hence analogical reasoning.

SYMBOLIC CONNECTIONISM

More basic than the issue of where in the brain the PSS is realized is the issue of what types of computations it employs to perform symbol manipulation. To address this issue, we have been developing a neural-network model of analogy called LISA (Learning and Inference with Schemas and Analogies). LISA represents an approach to building a PSS that we term symbolic connectionism (Hummel & Holyoak, 1997, in press; Holyoak & Hummel, in press). We have argued that one basic requirement for a PSS is the ability to represent roles (relations) independently of their fillers (arguments), which makes it possible to appreciate what different symbolic expressions have in common, and therefore to generalize flexibly from one to the other. In addition, to compose symbols into systematic structures—and to appreciate how those structures differ—it is necessary to explicitly bind relational roles to their fillers. "Jim loves Mary" differs from "Mary loves Jim", not in the representation of Jim, Mary, and loves, but in the binding of Jim and Mary to roles of the love relation. What gives a symbolic representation its power is precisely this capacity to represent roles independently of their fillers and at the same time to express the binding of roles to fillers dynamically—that is, without changing the representation of the roles or fillers (Fodor & Pylyshyn, 1988; Holyoak & Hummel, in press).

The symbolic connectionist framework that we have been developing seeks to realize these properties in neural networks. Traditional symbolic representations in cognitive science (generally in predicate-calculus-style notations) make no claim to be neurally plausible, as they permit arbitrary operations to create and move symbols freely from one structure to another. Early computational models of analogy, such as SME (Falkenhainer, Forbus & Gentner, 1989) and ACME (Holyoak & Thagard, 1989), were based on traditional symbolic representations, which render them inadequate as psychological and neural models (Hummel & Holyoak, 1997). Neural systems, which disallow such arbitrary operations, need some alternative means for composing invariant representations into symbolic structures—that is, for dynamically binding roles to their fillers.

Symbolic connectionist models (Holyoak & Hummel, in press; Hummel & Holyoak, 1997) and their precursors (Hummel & Biederman, 1992; von der Malsburg, 1981) use synchrony of firing for this purpose. The basic idea is that if two elements are bound together, then the neurons (or units in an artificial neural network) representing those elements fire in synchrony with one another; critically, elements that are not bound together fire out of synchrony. For example, to represent "Jim loves Mary", the units for Jim would fire in synchrony with the units for lover, while Mary fires in synchrony with beloved. To represent "Mary loves Jim", the very same units would be placed into the opposite synchrony relations, so that Mary fires in synchrony with lover while Jim fires in synchrony with beloved.

Symbolic connectionism represents a striking difference (and, we would argue, a striking advance) over traditional symbolic architectures of cognition (e.g., Anderson, 1993; Rosenbloom et al., 1991). One advantage of symbolic connectionism derives from an apparent weakness: It is hard to do symbol manipulation in a connectionist architecture. This is because symbol manipulation requires dynamic binding, and dynamic binding is difficult to perform in a connectionist architecture (Hummel & Stankiewicz, 1996, in press). In the case of dynamic binding by synchrony of firing, some mechanism has to get the right units into synchrony with one another and (what is even more difficult) keep them out of synchrony with all the other units. It takes work to establish synchrony and (especially) asynchrony, and some process must perform this work. A likely neural system for performing such operations is the human DLPFC.

In a traditional symbol architecture, by contrast, bindings are unlimited and require no special capabilities. Of course, a theorist may opt to impose some limit on binding, in deference to the glaring fact that people have limited capacity to make and break role bindings; but this will simply be an ad hoc "add on" rather than a deep implication of the proposed symbolic architecture. In contrast, a model that represents bindings with synchrony (e.g., LISA and related models such as JIM; Hummel & Biederman, 1992; Hummel & Stankiewicz, 1996), is inherently limited in the number of things it may simultaneously have active and mutually out of synchrony with one another (although there is no theoretical limit on the number of entities in any one synchronized group). That is, there is a limit on the number of distinct bindings such a model may have in working memory at any one time (Hummel & Holyoak, 1997; Shastri & Ajjanaggade, 1993). Humans, too, have limited working memory and attention. Symbolic connectionism—as an algorithmic theory of symbol systems—thus provides a natural account of the fact that humans have a limited working memory capacity.

We will now review the LISA model, and then consider recent psychological and neural evidence that human analogical reasoning is closely tied to working memory.

THE LISA MODEL

Analog Representation, Retrieval and Mapping

We will first sketch the LISA model and its approach to analog retrieval and mapping. These operations are described in detail (along with simulation results) by Hummel and Holyoak (1997). The core of LISA's architecture is a system for actively (i.e., dynamically) binding roles to their fillers in working memory (WM) and encoding those bindings in LTM. LISA uses synchrony of firing for dynamic binding in WM (Shastri & Ajjenagadde, 1993). Case roles and objects are represented in WM as distributed patterns of activation on a collection of semantic units (small circles in Figure 1); case roles and objects fire in synchrony when they are bound together and out of synchrony when they are not.

Every proposition is encoded in LTM by a hierarchy of structure units (see Figures 1 and 2). At the bottom of the hierarchy are predicate and object units. Each predicate unit locally codes one case role of one predicate. For example, love1 represents the first (agent) role of the predicate "love", and has bidirectional excitatory connections to all the semantic units representing that role (e.g., emotion1, strong1, posi-

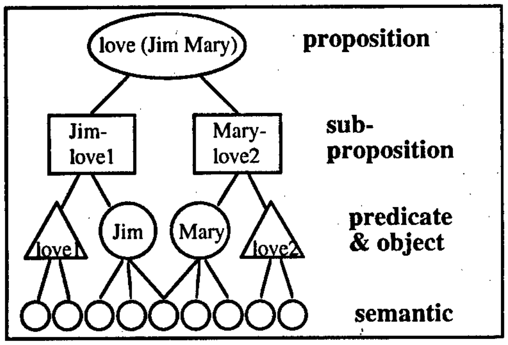


Figure 1. Illustration of the LISA representation of the proposition "love (Jim, Mary)".

tive1, etc.); love2 represents the patient role and is connected to the corresponding semantic units (e.g., emotion2, strong2, positive2, etc.). Semantically-related predicates share units in corresponding roles (e.g., love1 and like1 share many units), making the semantic similarity of different predicates explicit. Object units are just like predicate units except that they are connected to semantic units describing things rather than roles. For example, the object unit Mary might be connected to units for human, adult, female, etc., whereas rose might be connected to plant, flower, and fragrant.

Sub-proposition units (SPs) bind roles to objects in LTM. For example, "love (Jim, Mary)" would be represented by two SPs, one binding Jim to the agent of loving, and the other binding Mary to the patient role (Figure 1). The Jim+agent SP has bidirectional excitatory connections with Jim and love1, and the Mary+patient SP has connections with Mary and love2. Proposition (P) units reside at the top of the hierarchy and have bidirectional excitatory connections with the corresponding SP units. P units serve a dual role in hierarchical structures (such as "Sam knows that Jim loves Mary"), and behave differently according to whether they are currently serving as the "parent" of their own proposition or the "child" (i.e., argument) of another (Hummel & Holyoak, 1997). It is important to emphasize that structure units do not encode semantic content in any direct way. Rather, they serve only to store that content in LTM, and to generate (and respond to) the corresponding synchrony patterns on the semantic units.

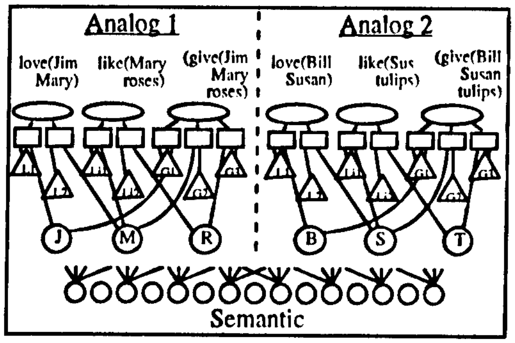


Figure 2. Representation of the "loves and flowers" analogy. Shapes (triangle, restangle, etc.) correspond to classes of units as in fig. 1. Not all connections are shown.

The final component of LISA's architecture is a set of mapping connections between structure units of the same type in different analogs. Every P unit in one analog shares a mapping connection with every P unit in every other analog; likewise,

SPs share connections across analogs, as do objects and predicates. For the purposes of mapping and retrieval, analogs are divided into two mutually exclusive sets: a driver and one or more recipients. Retrieval and mapping are controlled by the driver.

(There is no necessary linkage between the driver/recipient distinction and the more familiar source/target distinction.) LISA performs mapping as a form of guided pattern matching. As P units in the driver become active, they generate (via their SP, predicate and object units) patterns on the semantic units (one pattern for each role-argument binding). The semantic units are shared by all propositions, so the patterns generated by one proposition will activate one or more similar propositions in LTM (analogical access) or in WM (analogical mapping). Mapping differs from retrieval solely by the addition of the modifiable mapping connections. During mapping, the weights on the mapping connections grow larger when the units they link are active simultaneously, permiting LISA to learn the correspondences generated during retrieval. These connection weights also serve to constrain subsequent memory access. By the end of a simulation run, corresponding structure units will have large positive weights on their mapping connections, and non-corresponding units will have strongly negative weights.

Inference and Schema Induction

Augmented with intersection discovery and unsupervised learning, LISA's approach to mapping supports inference and schema induction as a natural extension (Hummel &Holyoak,

1996). Consider the previous “love and flowers” analogs (Figure 2). During mapping, corresponding elements in the two analogs will become active simultaneously. For instance, “love (Jim, Susan) will fire out of synchrony (Figure 3a). Jim shares male with Bill, and Mary shares female with Susan, so a natural proposition to induce from these correspondences is “loves (male, female)” (Figure 3b). To induce this part of the schema, it is necessary to (a) make explicit what corresponding elements have in common, and (b) encode those common elements into LTM as a new proposition.

LISA performs (a) by means of a simple type of intersection discovery. Although we have described the activation of semantic units only from the perspective of the driver, the recipient analog also feeds activation to the semantic units. The activation of a semantic unit is a linear function of its inputs, so any semantic unit that is common to both the driver and recipient will receive input from both and become roughly twice as active as any semantic unit receiving input from only one analog. Common semantic elements are thus tagged as such by their activation values.

These common elements are encoded into LTM by means of an unsupervised learning algorithm. In addition to structure units representing the known source and target analogs, LISA has a collection of unrecruited structure units (i.e., units with random connections to one another and to the semantic units) that reside together in a third “schema analog” (Figure 3). Unrecruited predicate and object units have input thresholds that only allow them to receive input from highly active semantic units — that is, semantic units that are common to both the driver and recipient analogs. Such semantic units are depicted in dark gray in Figure 3. Without the aid of an external teacher, these unrecruited schema units learn to respond to these common elements of the known analogs. Simultaneously, unrecruited SP units learn to respond to specific conjunctions of predicate, object, and (in the case of hierarchical propositions) P units, and unrecruited P units learn to respond to specific combinations of SP units. The result is that propositions describing thecommon elements of the known analogs are encoded into LTM as a third analog — a schema. Figure 3 illustrates this process for one proposition in the “love and flowers” analogy.

LISA accomplishes analogical inference by the same unsupervised learning algorithm as

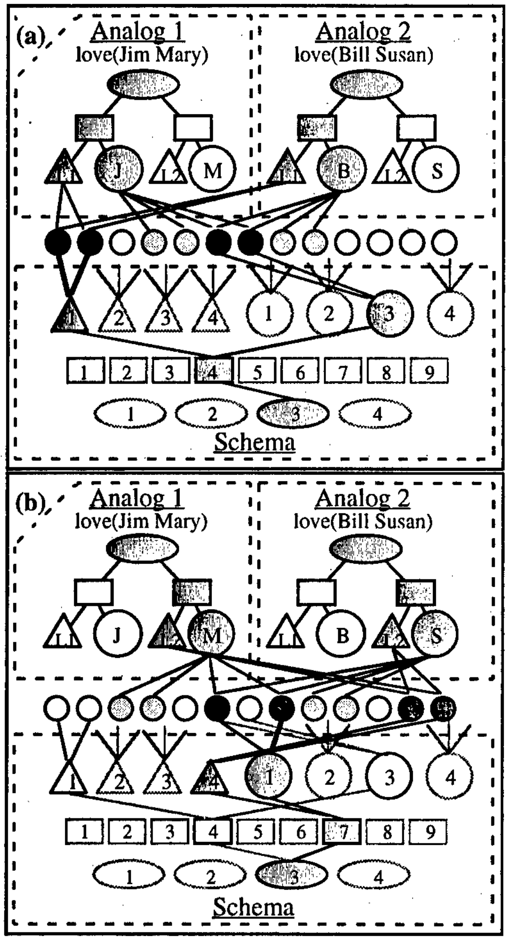


Figure 3. Jim+love-agent in Analog1 activates Bill+love-agent in Analog 2. In the Schema, predicate unit 1 is recruited for love agent, and object unit 3 is recruited for the intersection of Jim and Bill ("human" and "male"). SP 4 is recruited for human male (object 3) bound to love agent (pericate 1). Propositi on unit 3 begins to be recruited. (b) Mary+love-patient in Analog 1 activates Susan+love-patient in Analog 2. Predicate 4 is recruited fro love-patient; object 1 is recruited for "human" and "female". SP 7 is recruited for the binding of predicate 4 and object 1. Propsition unit 3 now codes "love(human male, human female)".

used for schema induction, except that the unrecruited units reside not in a completely separate analog (the to-be-induced schema), but in the target itself.

WORKING MEMORY AND RELATIONAL REASONING

Grouping Effects and Mapping Asymmetries

A key distinction between LISA and previous computational models is its emphasis on the role of working memory in controlling mapping. In LISA, mapping is a directional, capacity-limited and sequential process. The directional aspect of mapping follows from the driver/recipient distinction. If a driver analog contains more propositions than WM can hold, the propositions must be fired in small groups (roughly, up to six role bindings, or 2-3 propositions, at a time).

The role of WM in LISA's operation leads to predictions about the influence of grouping propositions on the performance of the model (and hence people). For example, if the text of the driver analog is thematically connected (e.g., by causal relations), then mapping may be more accurate than if the text consists of causally unrelated propositions. This prediction was confirmed in a study by Keane (1997).

Our group (Kubose, Holyoak & Hummel, 1997) extended Keane's procedure to demonstrate that the impact of causal structure on mapping is inherently asymmetrical. Although models such as as SME (Bowdle & Gentner, 1997) and ACME (Holyoak, Novick & Melz, 1994) can account for asymmetries that arise in post-mapping analogical inferences, only LISA and the IAM model (Keane, Ledgeway & Duff, 1994) predict asymmetries in the mapping stage itself (and only LISA predicts asymmetries as measured by mapping accuracy). In LISA, the driver but not the recipient is processed sequentially, and hence it is the driver that is sensitive to groupings of propositions. It follows that mapping performance with isomorphic analogs will be more accurate if the driver analog is causally connected and the recipient analog is not, rather than vice versa.

Kubose et al. manipulated the driver/recipient status by having subjects first answer questions about one or the other analog, and then asking directed mapping questions (i.e., for each object and relation in the driver, subjects were asked to provide the corresponding element from the recipient). The results supported LISA's prediction, indicating that mapping performance was more accurate when the driver analog, rather than the recipient, had causal content.

Other experiments supported LISA's interpretation of causal effects on mapping as being mediated by selective grouping of propositions. If neither analog was thematic, but certain propositions in the driver were optimally grouped simply by drawing a box around them and asking subjects to consider them together, mapping accuracy was improved.

Other recent experiments by our group (Grewall, Law & Holyoak, in progress) have tested a different type of prediction that LISA makes about the role of working memory in mapping. In accord with the theory of relational complexity developed by Halford and his colleagues (Halford & Wilson, 1980; Halford, Wilson & Phillips, in press), LISA predicts that the complexity of mappings is constrained by the availability of WM resources to maintain multiple dynamic role bindings concurrently. It follows that if WM is restricted by adding dual-task requirements (e.g., digit memory load), which are known to compete for WM capacity (e.g., Hitch & Baddeley, 1976; Gilhooly et al., 1993), the ability to make relationally complex mappings will be impaired.

In order to determine if a dual task will shift the preferred basis for making comparisons, we asked subjects to map a set of stimuli with ambiguous mapping. These stimuli, created by Markman and Gentner (1993), are pairs of pictures (e.g., a man bringing groceries to a woman; a woman feeding nuts to a squirrel) in which one element of the first picture (e.g., the woman) can map to either of two elements in the second (the woman, on the basis of perceptual similarity, or the squirrel, based on the shared role of recipient-of-food). We found that adding a dual task (concurrent digit load) caused a shift from relational to more direct perceptual similarity as the basis of mappings. Such a shift is predicted by LISA because finding the relational match is more dependent on WM resources, which are reduced by a concurrent memory load.

In addition, Tohill and Holyoak (in progress) have shown similar reductions in relational matches when subjects' anxiety level is increased prior to the mapping task (by a difficult backwards-counting task). The detrimental impact of anxiety on relational mapping is consistent with theories of anxiety that emphasize its restrictive impact on WM resources (Eysenck & Calvo, 1992).

Neuropsychological and Neuroimaging Studies

We have also begun to investigate the neural locus of the operations that support relational reasoning. Investigations by our group have revealed selective deficits in relational processing in tasks similar to analogy, such as simple variants of Raven's Progressive Matrices problems (see Carpenter, Just & Shell, 1990), for patients with focal degeneration of the prefrontal cortex (Waltz et al., in press). The patients tested were diagnosed with frontotemporal dementia (FTD), a dementing syndrome resulting in the degeneration of anterior regions of cortex (Brun et al., 1994). In the early stages of FTD, the degenerative process tends to be localized to either prefrontal or anterior temporal cortical areas, with eventual involvement throughout all cortical regions in advanced stages. This makes possible the division of patients with mild FTD into two subgroups of patients. In the frontal variant of FTD, damage is initially localized in prefrontal cortex. Patients with the temporal variant of FTD often exhibit semantic dementia, characterized by impairments in semantic knowledge (Graham & Hodges, 1997).

Waltz et al. (in press) found that, relative to patients with damage to anterior temporal cortex, patients with degeneration of prefrontal cortex show dramatic impairment in the ability to make inferences requiring the integration of multiple relational representations. For example, performance on a set of matrix problems showed striking differences between patients with damage to prefrontal cortex and those with damage to anterior temporal cortex and normal controls in the ability to integrate multiple relational premises. The two patient groups did not differ either from each other or from normals in the average proportion of correct responses given to problems not requiring relational integration (i.e., problems with variation on at most one dimension). However, on problems that required integration (those with variations on two dimensions), the patients with prefrontal cortical damage were catastrophically impaired compared to patients with anterior temporal lobe damage as well as normal controls.

To complement the neuropsychological studies, a number of researchers in our group at UCLA (Kroger, Holyoak, Bookheimer & Cohen; see Kroger, 1998) have begun to perform neuroimaging studies to investigate the neural basis of relational processing in normal college students. Previous functional imaging studies of reasoning have shown involvement of the same areas of cortex as are activated in working-memory tasks, especially DLPFC (e.g., Prabhakaran et al., 1997), but have not systematically manipulated relational complexity.

We have constructed materials matched closely in terms of visuospatial attributes, but varying in relational complexity (Halford & Wilson, 1980; Halford et al., in press). A pilot experiment in progress uses variants of Raven's Progressive Matrices problems which vary the number of relational that that must be considered in the production of an inductive inference. These problems are more complex versions of the matrix problems used with FTD patients by Waltz et al. (in press), suitable for use with normal college students. In pilot work in progess, we are using five levels of relational complexity. Behavioral data show increasing reaction times as relational complexity increases, confirming that we are tapping into increasing complex cognitive processes. Initial analyses of data from the first subject to be tested reveal that activation in prefrontal cortex (but not parietal cortex) increases monotonically with relational complexity (Kroger, 1998).

These neuropsychological and initial neuroimaging results provide support for our hypothesis that relational processing may form the core of an executive component of prefrontal working memory, which implies both the active maintenance of information and its processing. In other words, relational integration—and specifically, dynamic variable binding—may be the "work" done by working memory. We have recently begun to simulate our neuropsychological findings using the LISA model (Holyoak et al., 1998; Hummel et al., 1998).

CONCLUSION

Symbolic connectionism, as instantiated in models such as LISA, offers a possible account of the general form of the Physical Symbol System that underlies human (and other primate) relational reasoning. LISA provides a solution to the problem (forcefully posed by Fodor & Pylyshyn, 1988) of representing knowledge over a distributed set of units while preserving systematic relational structure. Like previous models based on traditional symbolic representations, LISA is able to retrieve and map analogs based in large part on structural constraints. But in addition, LISA is able to capitalize on its distributed representations of meaning to integrate analogical mapping with a flexible mechanism for analogical inference and schema induction.

A key aspect of LISA, given its use of dynamic binding, is that analogical processing (and relational reasoning in general) is heavily constrained by working-memory resources. In order to make relationally complex mappings, the reasoner must be able to consider multiple role bindings together. We can now begin to see not only what mappings are "natural" for human reasoners, but also how they may be computed in neural systems, and what regions of the brain are necessary for performing these computations.

ACKNOWLEDGEMENT

This work was supported by NSF Grant SBR-9729023.

REFERENCES

Anderson, J. R. (1993). Rules of the mind. Hillsdale, NJ: Erlbaum.

Benson, D. F. (1993). Prefrontal abilities. Behavioral Neurology, 6, 75-81.

Bowdle, B. F., & Gentner, D. (1997). Infomativity and asymmetry in comparisons. Cognitive Psychology, 34, 244-286.

Brun, A., Englund, B., Gustafson, L., et al. (1994). Clinical and neuropathological criteria for frontotemporal dementia. Journal of Neurology, Neurosurgery and Psychiatry, 57, 416-418.

Carpenter, P.A., Just, M.A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. Psychological Review, 97, 404-431.

Eysenck, W. E., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. Cognition and Emotion, 6, 409-434.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. Artificial Intelligence, 41, 1-63.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. In S. Pinker & J. Mehler (Eds.), Connections and symbols (pp. 3-71). Cambridge, MA: MIT Press.

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. Memory & Cognition, 21, 115-124.

Gillan, D. J., Premack, D., & Woodruff, G. (1981). Reasoning in the chimpanzee: I. Analogical reasoning. Journal of Experimental Psychology: Animal Behavior Processes, 7, 1-17.

Grafman, J., Holyoak, K. J., & Boller, F. (Eds.) (1995). Structure and functions of the human prefrontal cortex. New York: New York Academy Sciences.

Graham, K.S., & Hodges, J.R. (1997). Differentiating the roles of the hippocampal complex and the neocortex in long-term memory storage: Evidence from the study of semantic dementia and Alzheimer's disease. Neuropsychology, 11, 77-89.

Halford, G. S., & Wilson, W. H. (1980). A category theory approach to cognitive development. Cognitive Psychology, 12, 356-411.

Halford, G. S., Wilson, W. H., & Phillips, S. (in press). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. Brain and Behavioral Sciences.

Hitch, G. J., & Baddeley, A. D. (1976). Verbal reasoning and working memory. Quarterly Journal of Experimental Psychology, 28, 603-621.

Holyoak, K. J., & Hummel, J. E. (in press). The proper treatment of symbols in a connectionist architecture. In E. Dietrich & A. Markman (Eds.), Cognitive dynamics: Conceptual change in humans and machines. Cambridge, MA: MIT Press.

Holyoak, K. J., Novick, L. R., & Melz, E. R. (1994). Component processes in analogical transfer: Mapping, pattern completion, and adaptation. In K. J. Holyoak & J. A. Barnden (Eds.), Advances in connectionist and neural computation theory, Vol. 2: Analogical connections (pp. 130-180). Norwood, NJ: Ablex.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. Cognitive Science, 13, 295-355.

Holyoak, K. J., & Thagard, P. (1995). Mental leaps: Analogy in creative thought. Cambridge, MA: MIT Press.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. Psychological Review, 99, 480-517.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. Psychological Review, 104, 427-466.

Hummel, J. E., & Holyoak, K. J. (in press). From analogy to schema induction in a structure-sensitive connectionist model. In T. Dartnall & D. Peterson (Eds.), Creativity and computation. Cambridge, MA: MIT Press.

Hummel, J. E., & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McClelland (Eds.), Attention and performance XVI: Information integration in perception and communication (pp. 93-121). Cambridge, MA: MIT Press.

Hummel, J. E., & Stankiewicz, B. J. (in press). Two roles for attention in shape perception: A structural description model of visual scrutiny. Visual Cognition.

Hummel, J. E., Waltz, J. A., Knowlton, B. J., & Holyoak, K. J. (1998). A symbolic connectionist model of the impact of prefrontal damage on transitive reasoning. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society.

Keane, M. T. (1997). What makes an analogy difficult? IAM predicts the effects of order and causal relations on analogical mapping. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 1-22.

Keane, M. T., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. Cognitive Science, 18, 387-438.

Kroger, J. K. (1998). Human processing of relationally complex representations: Cognitive and neural components. Ph.D. dissertation, Department of Psychology, UCLA.

Kubose, T. T., Holyoak, K. J., & Hummel, J. E. (1997). Asymmetries in analogical mapping: A test of a process model. In M. G. Shafto & P. Langley (Eds.), Proceedings of the Nineteenth Conference of the Cognitive Science Society (p. 976). Hillsdale, NJ: Erlbaum.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. Cognitive Psychology, 23, 431-467.

Marr, D. (1980). Vision. Freeman: San Francisco.

Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G., & Gabrieli, J. D. E. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. Cognitive Psychology, 33, 43-63.

Premack, D. (1983). The codes of man and beasts. Behavioral and Brain Sciences, 6, 125-167.

Robin, N., & Holyoak, K. J. (1995). Relational complexity and the functions of prefrontal cortex. In M. S. Gazzaniga (Ed.), The cognitive neurosciences (pp. 987-997). Cambridge, MA: MIT Press.

Rosenbloom, P. S., Laird, J. E., Newell, A., & McCarl, R. (1991). A preliminary analysis of the Soar architecture as a basis for general intelligence. Artificial Intelligence, 47, 289-325.

Newell, A. (1980). Physical symbol systems. Cognitive Science, 4, 135-183.

Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

Shallice, T., & and Burgess, P. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), Frontal lobe function and dysfunction (pp. 125-138). New York, NY: Oxford University Press.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. Behavioral and Brain Sciences, 16, 417-494.

Tomasello, M., & Call, J. (1997). Primate cognition. New York: Oxford University Press.

Vera, A. H., & Simon, H. A. (1993). Situated action: A symbolic interpretation. Cognitive Science, 17, 7-48.

Vera, A. H., & Simon, H. A. (1994). Reply to Touretzky and Pomerleau: Reconstructing physical symbol systems. Cognitive Science, 18, 355-360.

von der Malsburg, C. (1981). The correlation theory of brain function. Internal Report 81-2, Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., Thomas, C. R., & Miller, B. L. (in press). A system for relational reasoning in human prefrontal cortex. Psychological Science.

THE ROLE OF SIMILARITY IN HOW WE CATEGORIZE THE WORLD

James A. Hampton
Department of Psychology
City University
j.a.hampton@city.ac.uk

Given the key importance of the concept of "similarity" for understanding analogy, the purpose of my paper will be to investigate a parallel issue - the role of similarity in understanding categorization.

It may seem almost tautological to say that we categorize the world into categories of similar objects, persons or events. Similarity is after all merely an extension of the notion of "sameness". Similarity may just be sameness in respect of a particular set of features or dimensions. So I may be similar to a colleague in working for the same organization, having the same job title, or having the same number of children. Similarity may also be **closeness** on a continuous dimension, so that I and my colleague may share a similar colour of hair, a similar salary or a similar personality.

As these examples quickly illustrate, while we expect categories to be composed of similar elements, there is a major difficulty in **explaining** categorization in terms of raw similarity defined as sameness or closeness on a set of dimensions. The problem is that there is an indefinitely large number of such dimensions, and there could therefore be any number of reasons for placing two items in the same category and any number of reasons for placing them in different categories.

The idea that we classify together those things that we find similar has had a chequered history in psychology. While there was considerable theoretical and empirical interest in the development of similarity-based classification models in the 1970s, particularly with Rosch and Mervis' prototype theory, and Medin & Schaffer's Exemplar model, (Medin & Shaffer, 1978; Rosch, 1975), subsequently the field has split into two very distinct camps. On the one hand increasingly sophisticated computational models have been developed to explain how people learn classifications on the basis of similarity. Most notable in this area are developments of exemplar storage models based on Medin and Shaffer's context model. The models assume that we encode stimuli in a multi-dimensional similarity space, and learn classifications through one of a number of possible algorithms. In Nosofsky's Generalized Context Model (Nosofsky, 1988) similarity of a new stimulus is computed to all the stored exemplars of each category that has been learned, and a choice rule determines the likelihood of classification in a particular category. In Ashby and Gott's (1988) Decision Bound approach, the space is divided up by hyperplanes that delimit the boundaries where the probability of belonging in one category equals that of belonging in its neighbour. These different models have been shown to provide an excellent fit to a range of experimental data in classification and recognition tasks.

Meanwhile, researchers in higher level cognition have questioned the degree to which the notion of similarity is sufficiently clearly defined and well enough constrained to serve as an explanation of how we actually carve up and categorize the real world around us, as opposed to the artificial stimulus worlds dreamt up by psychologists devising their experiments. In particular there is the major concern of finding an independently motivated account of why we attend to particular dimensions of our environment rather than others. Similarity-based categorization can only be made to work given a specification of relevant dimensions. It is perfectly possible for dimensional weights to be adapted to the distribution of stimuli in order to maximise the coherence of categories in the space, and there is evidence that this does happen. But the selection of dimensions from which to start is a far from trivial issue.

In this talk, I will discuss arguments and review evidence for and against basing categorization on similarity, and conclude that, construed broadly, similarity may still have a key role to play in explaining how most of our conceptual categories function.

SIMILARITY-BASED CATEGORIZATION

What is the evidence that similarity plays a role in categorization? To answer this question we need to be quite precise about what we mean by similarity. We form categories of many different kinds in the course of everyday cognition, and it could be claimed that they are *all* based on similarity. But this would be to render the notion so broad as to be empty or more probably circular.

To begin with examples of categories that are **not** good candidates for a similarity-based account, Barsalou (1983) pointed to the existence of what he termed *ad hoc* categories such as Birthday Presents for Your Mother, or Things to Take on a Camping Trip. Members of these categories are of course similar in one important respect — things to take on a camping trip are all similar in as much as they are all good things to have along when camping. But this tautological similarity does not go far in explaining how this category is constructed. Nor does it appear that the degree to which something is a good member of the category is related in any way to its similarity to other members in any respect *other than* its property of being in the category.

Another class of categories which could only tautologically be explained in terms of similarity is the class of concepts with *explicit* definitions. Thus belonging to the conceptual category of Triangle depends on a small number of explicit criteria, such that only similarity *in those respects* is relevant to class membership. To say that all triangles are similar to each other in respect of having three straight sides, three angles, and internal angles that sum to 180° is to say little more than that all triangles possess all these properties. At the same time, the ratios of the three sides or the three angles may affect perceived similarity of actual triangles, but are clearly of no relevance to the issue of category membership. Thus similarity reduces to identity in certain restricted respects, while other respects are treated as totally irrelevant. Categories of this kind are clearly *not* based on similarity, except in a purely tautological sense. Similarity must mean more than simple identity on a particular set of dimensions, and there should be some independent justification for treating otherwise salient dimensions as being irrelevant to categorization.

By contrast, we form many other categories, many of them stable and long-term parts of our conceptual repertoire, which *do* show a strong *prima facie* link to similarity. These categories are characterized by having *no explicit definition* (unlike ad hoc categories or explicitly defined categories), a number of associated properties which are *generally* true of category members, although not universally so, and a graded structure such that some items are more clearly and uncontroversially members of the category than are others. Rosch and Mervis (1975) termed these concepts "family resemblance" or Prototype Concepts. Prototypes are ideal or central tendencies around which categories form. The category is then composed of all items that are sufficiently similar to the prototype (for a formal treatment see Hampton, 1995a). Prototype theory answers the key question of how dimensions are selected by proposing that our biological inheritance and social and cultural environment provide the dimensions along which we note similarity and difference. Where a number of these dimensions correlate in our experience, then a category of similar items is formed, to which we give a name, and which we can then use as a concept in our thinking and language. Once the dimensions have been determined, clustering of the world into classes is relatively automatic. Indeed there are advanced statistical theories of how items may be clustered based on partially correlated dimensions (van Mechelen et al., 1993).

There are several iterative feedback loops in this process. For an individual learning the categories of his or her culture, the first attempts to understand the relevant dimensions may be incorrect and may need refinement through error correction. Keil and Batterman's (1984) study of the Characteristic-to-Defining shift in young children shows just this type of effect. Younger children took account of more perceptually striking dimensions in making categorization judgements about concepts such as Island, Uncle or Lunch, while the older children had homed in on the correct concepts as determined by adult usage of the words.

At the cultural level, in order to obtain a cleaner and more generally useful set of categories, the weights of dimensions get adjusted or new dimensions are constructed as concepts evolve. The reason that younger children have to adapt their concepts to pick up these more hidden or subtle conceptual distinctions is that to suit its purposes our culture has developed concepts based on a deeper level of structure containing more relational information and less dependent on mere appearance.

It is at this point in the story that a number of psychologists have argued that something other than mere similarity and feature weights must be playing a role. Part of our drive for knowledge and understanding is the search to replace similarity-based clusters based on perceptual appearance by explicitly defined concepts with broad explanatory power. Keil (1989) refers to this as the principle of "original sim" — that children's initial concepts are based on pure similarity, which is then replaced in time with deeper, more theory-like kinds of conceptual understanding.

A paradigm example of this process can be seen in the progress of medical science. When medical research first tackles a phenomenon it defines a *syndrome* — a cluster of symptoms, and conditions of occurrence, with some predictive value in terms of treatment and prognosis. (Most mental illnesses are at this stage of understanding.) It is characteristic of syndromes that cases may be more or less typical, and more or less clear members of the syndrome. Frequently cases may arise that are borderline to the syndrome, possessing some similarity to typical cases, but not enough to be clearly identifiable as an example. Discovery of an aetiology linked to the syndrome — such as an infectious organism, a genetic marker, or an identifiable biochemical malfunction — will usually allow the syndrome to be replaced by a clearly defined disease or condition category, with its own set of diagnostic tests. Note that the set of patients and their symptoms has not changed — the world has not become more clear-cut in any way. However whereas before a case was borderline because it showed marginal levels of similarity to other cases, a case will now be borderline if the critical diagnostic tests do not come out with a clear answer. There is a shift from an uncertainty which is *conceptual* in its origin, to an uncertainty which is *epistemological* — that is to say that a case is now borderline because we cannot discover clearly enough whether the defining agent is at work. Our uncertainty has to do with our state of knowledge in the particular case, rather than our state of understanding of such cases in general.

This extended analogy with medical science serves as a template for the debate that followed publication of Murphy and Medin's (1985) attack on similarity as a basis for natural concepts. Physicians seek to *explain* the presenting symptoms through a causal account. In an analogous fashion, Murphy and Medin argued that we use our concepts as ways of explaining the world to ourselves and others. To take one of their examples, if we see someone jump fully clothed into a swimming pool at a party, we may categorize them as *drunk*. We do not have to do this by comparing their behaviour to similar examples of drunken behaviour that we have seen in the past (although actually this *might* be how we do it), but according to Murphy and Medin, we can make the categorization by looking for the category that best provides an *explanatory account* of the behaviour that we are seeing. Such a process for categorizing through causal or explanatory "mini-theories" is a much more powerful means of categorizing, as it is possible to use it to categorize examples that are far removed from any familiar experiences that we may have had in the past. According to this account, the dimensions on which we categorize are themselves determined by a deeper causal explanatory theory which links the observable facts to a deeper underlying cause, and so makes the whole category a coherent set. The crucial point that Murphy and Medin make is that to determine that a particular drunken behaviour is **similar to** other examples of drunken behaviour seen previously requires that we can specify in just what respects that similarity is measured. But the only way to do this is to have a theory of what effect alcohol has generally on behaviour. The determination of similarity depends on the theory, and so cannot itself play an explanatory role in the categorization.

It follows from this critique that we categorize not on the basis of a similarity cluster (akin to a syndrome), but on the basis of selecting the concept that best explains the instance to be categorized (as in a disease category). This alternative account of categorization has also had wide acceptance in the developmental field (Keil, 1989).

The difference between similarity and explanation-based or "causal theory" accounts of categorization was brought into sharp focus in a paper by Rips (1989). Rips attacked the unconstrained nature of similarity as a basis for categorization, and reported a number of demonstrations of cases where the similarity account clearly fails. Each of these demonstrations involved the discovery of a non-monotonic dissociation in the relation between similarity and categorization. If categories are formed around prototypes, then it should not be the case that one item could be more similar (or more typical) of the category than another, but yet less likely to belong. In formal terms, this means that there should be a monotonic function relating similarity to a category and membership in that category. Rips provided three cases where this constraint was broken.

In his first case, subjects were asked to consider a hypothetical item that was exactly half way between two categories, one a fixed category and the other a variable category. For example they had to imagine an object that was half way between the largest US quarter they had seen and the smallest pizza they had seen. Subjects then judged whether this object was either (a) more similar to or typical of one category rather than the other, or (b) more likely to be a member of one category rather than the other. Rips reported a dissociation between similarity and typicality on the one hand, where people generally considered similarity to be about equal to each category, and likelihood of membership on the other hand, where people generally judged the object more likely to be in the variable category (the pizza in this case). Since similarity to the two categories was equal, but categorization was strongly biased in favour of one, Rips argued that categorization behaviour was dissociated from similarity.

Rips' second example involved a creature (or artifact) which metamorphosed into something else. For example a bird-like creature was transformed into an insect-like creature through an environmental accident. When asked whether it was more similar to or more typical of a bird as opposed to an insect, people went for the insect category. However when asked which type of creature it was more likely to be, they judged the creature (marginally) more likely to be a bird. Once again there was a dissociation in that whereas similarity pointed to categorization in one category (insect), actual categorization preferences were for placing the creature in the other (bird).

The third example was reported in a paper by Rips and Collins (1993). Subjects were given information about the shapes of two (non-normal) distributions of values on some dimension - for example daily maximum temperatures for two particular locations. They were then given particular values and asked to judge their typicality as an example of each distribution, or asked to say which distribution the item was more likely to belong to. Under these conditions, people tended to base similarity judgments on distance from some measure of central tendency. Likelihood of categorization however was based on a more extensional form of reasoning, employing intuitive statistical reasoning to find the more likely category.

There is no space in this paper to go into a detailed discussion of the validity of Rips' three cases of non-monotonicity (but see Hampton, 1997, for a fuller discussion). What is clear is that dissociations between typicality and category membership can be demonstrated albeit with relatively non-standard types of material. The first case asked people to imagine an object which is specified *only* by its size. The second involved a creature whose appearance changed, but about whose internal organs and genetic make-up subjects were told nothing, and the third case involved presenting subjects with strong cues to employ extensional reasoning using relative frequencies in their category judgments. (Physicians are familiar with the phenomenon of cases that may resemble condition A more than condition B, but where the extreme rareness of condition A means that a diagnosis of condition B is more likely to be correct.)

One aspect that all three demonstrations share is a presupposition that categorization is in fact all-or-none. Thus the object was either a coin or a pizza, it was either a bird or an insect, and either from one distribution or the other. The categorization task was always presented to the subject as one in which the *correct* categorization had to be *predicted* on the basis of the available information. As noted earlier, this presupposition is antithetical to the similarity-based approach where the correctness of a categorization is not something that can always be resolved. Some items are by their nature borderline to a class, and no further exploration would reveal their true nature any better.

EVIDENCE FOR SIMILARITY IN CATEGORIZATION

In the light of these various critiques of similarity-based categorization it is worth briefly reviewing the evidence *for* the prototype model. First there is the *fuzziness* of many of our concepts. When asked to reflect on the meaning of words like "fish", "art", or "sport", people find it very hard to give a theoretically satisfactory account of the underlying concepts. They are however very good at generating ways in which members of the category differ from other things in the same domain. They can also quickly recall or create examples to illustrate what a typical category member might be. There is apparently a rich source of semantic information associated with the concept, but it does not appear to be organized in anything like the neat structures proposed by the opponents of prototype theory. The lack of organization and internal coherence becomes particularly clear when people's reasoning with concepts has been studied. Hampton (1982) showed that people may quite willingly agree (for example) that School Furniture is a type of Furniture, and that a blackboard is a type of School Furniture, but yet disallow that a blackboard is a type of Furniture. Categorization was not treated as a universally transitive relation, in contradiction of both classical and even fuzzy logic (Zadeh, 1965). Instead, I argued that each separate category judgment was made on the basis of similarity. As the basis on which similarity changes between the two judgments, it is then quite possible to obtain intransitive categorizations.

Tversky and Kahneman (1983) found similar effects on subjective probability judgments. They found that people used similarity to prototype as a means of judging subjective likelihood, even when this strategy produced clearly illogical results, such as judging it more likely that a radical female student would have become a feminist bank teller, than that she would simply have become a bank teller. This conjunction fallacy was paralleled by the finding of overextension of conjunctive categories by Hampton (1988). People were willing to say for example that Chess was a Sport which is a Game, even though they had earlier judged that Chess was not a Sport. Hampton (1996a) replicated this result with a between-subjects design, and extended the demonstration of inconsistent classification to the case of negation. For example 80% of participants in one group considered Tree Houses to be Buildings, yet 100% of participants in another group considered them to be Dwellings that are *not* Buildings. Our conceptual categories display a degree of flexibility and context sensitivity which is much more easily captured by a similarity-based process than by a fixed theoretical schema. A recent study by Sloman (1997) is a further demonstration of how similarity can be shown to affect people's reasoning. In one demonstration, Sloman found that people were more likely to accept the truth of a logically necessary conclusion when the two premises were similar than when they were not. Similarity apparently pervades people's attempts to reason logically, and a very simple explanation for this finding is that our conceptual system is heavily dependent on similarity-based conceptual processes.

A critical test of similarity-based categorization is the extent to which categorization can be influenced by "irrelevant" kinds of similarity. There is a distinction in the literature, originally introduced by Smith, Shoben and Rips (1974), between Defining and Characteristic Features. It was their notion that there were many properties of objects which might determine how typical they were of their class, but which would be irrelevant to their category membership. Their example was that the ability to fly is very typical of birds, and so flying birds are more typical members of their class. Flight as such however is irrelevant to determining whether a creature is a bird or not, since there are both birds that do not fly and other creatures (notably insects) that do fly. Smith et al. termed this idea the Characteristic Feature Hypothesis. Hampton (1995b) set out to test whether Characteristic Features (CF) are in fact always irrelevant to categorization in practice. To test this idea, I created sets of six hypothetical objects for each of a number of concepts. Each object either possessed or lacked a full set of CF. In addition each object either had a full set of Defining Features (DF+), lacked at least one Defining Feature [DF-), or had a *partial match* to the Defining Features [DF?]. The aim of the experiment was first to show that when the object possessed the DF, categorization would be clearly positive, and when it lacked at least one DF, then it would be clearly negative, regardless of the CF. The critical test was then to be whether the CF would affect categorization when the DF were only partially matched. For example consider an object which *partially* matched the DF of umbrellas - it was designed to keep things from falling on you, but instead of protecting you from the rain it was intended to protect you from acorns and twigs when picnicking under a tree. Would this odd object be more likely to be categorized as an umbrella if it had the classical domed shape and material of umbrellas, than if it was built in some different shape and material?

In the event this critical second test could not easily be performed. The reason was that it proved very hard (even after four replications of the experiment with improved materials and improved instructions), to find CF which did not still influence categorization, even when the DF were clearly present or absent. For example one example of DF+, CF- was the following description:

"The offspring of two zebras, this creature was given a special experimental nutritional diet during development. It now looks and behaves just like a horse, with a uniform brown color."

When asked if this was really a zebra, only a third of the subjects agreed, the rest ignoring the genotype in favor of the phenotype, contrary to the assumptions of both biological theory and psychological essentialism. Similar problems occurred when I attempted to pit the intended function of artifacts (assumed to reflect their real nature) against their outward appearance. People tended to be influenced by similarity along dimensions which logical analysis suggests should be irrelevant — *unless* of course categorization is based on similarity calculated across a wide range of dimensions.

Returning to the critique offered by Rips (1989), an unpublished study by Hampton & Estes attempted partially to replicate Rips' transformation study. We felt that the design of the original study may have encouraged subjects to dissociate the similarity/typicality and categorization judgments, simply because both questions were always asked together after every scenario. We modified the procedure in a number of ways, the main one of which was to have different groups of students making judgments of typicality or judgments of categorization. The startling finding was that the dissociation completely disappeared. There were no differences in the mean ratings for typicality or categorization in any of the conditions. Thus when the creature was not yet transformed it was uniformly rated as typical of, and likely to belong in, the initial category. After the transformation, both typicality and categorization switched to the final category. When the nature of the transformation was changed from an "accidental" change induced by environmental pollution, to a "natural" change due to biological maturation, both typicality and categorization judgments showed some degree of switch towards the final category, but there was still no dissociation.

In a further unpublished study by Hampton & Hainitz, using a similar design, we varied whether the transformation affected just the surface external appearance (through surgical intervention) or affected the deeper internal biology of the creature (through environmental pollution). The degree to which the creature was believed to have changed was greater for the deep transformation than for the surface one, and there was a greater shift towards the final category for the typicality judgements than for categorization. Yet there was still no evidence of a clean dissociation in the two judgments.

DISSOCIATING CATEGORIZATION AND SIMILARITY IN NATURAL CATEGORIES

According to the Prototype Model, categorization proceeds by assessing the similarity of an instance or subclass to the concept prototype, and then testing whether it passes some threshold criterion for category membership. If this model is inadequate, then as Rips (1989) argued, it should be possible to demonstrate non-monotonicity between measures such as typicality or similarity to prototype (on the one hand) and likelihood of category membership (on the other). Hampton (1997) set out to discover to what extent non-monotonicity of this kind could be found in everyday common semantic categories. Rips (1989) used a variety of unusual examples to dissociate similarity and categorization, and it is questionable how generalizable such results are to the more usual process of deciding if subclass A is a member of category B. It is therefore interesting to know whether categorization in a common category such as Fish or Vehicle follows typicality in the category, or whether dissociations between the measures can be found. To answer this question, I reanalyzed a data set published in 1978 by McCloskey and Glucksberg, in which they had two groups of subjects making judgments about 18 semantic categories. One group were asked to make typicality judgments for a list of 30 items for each category, ranging from clear category members to clear non-members. A second group gave a simple Yes/No categorization decision about each item for each category. This second group returned a month later and made their categorization decisions a second time. McCloskey and Glucksberg (1978) found that the categorizations showed fuzziness in two respects. First, there was considerable disagreement amongst people over which items should be included in the categories and which should not. This disagreement was reflected in a large number of items with Categorization Probability at intermediate levels between 0 and 1. Second, there was a considerable degree of within-subject inconsistency when the follow-up test was made. High levels of disagreement and inconsistency were most noticeable for items in the *middle* of the typicality scale — that is for items that were neither clear members nor clear non-members. McCloskey and Glucksberg concluded that categorization in many semantic categories is fuzzy, rather than all-or-none, and that there is a considerable amount of instability in how we categorize.

The data from this research were published as an Appendix, and provided an opportunity to test for non-monotonicity directly. Typicality ratings are *prima facie* direct measures of how similar an instance or class is to the category prototype. The instructions for typicality emphasize that a high rating should be given to items that are *representative* or *good examples* of the class as a whole. On the other hand Categorization Probability is a simple way of measuring the degree to which something is categorized in a class. If we assume that there are random and individual sources of variation in categorization, then the group measure of how many subjects say X is in category Y may be taken as a fairly direct measure of the degree to which X is considered to belong in Y by each individual.

The data were therefore analyzed in order to examine the mathematical relationship between mean rated typicality and categorization probability. Technical details can be found in Hampton (1996b). The first conclusion was that there were clear differences between individual categories in terms of how clearly categorization probability could be predicted from typicality. For example, Sport showed a clear threshold function, with practically no systematic deviation from the expected pattern of categorization probability rising with typicality. For Fish on the other hand, there was a considerable spread of items above and below the threshold function, and plenty of evidence for non-monotonicity. There was no link however between how well the measures correlated and the kind of semantic domain. There were good and bad fits in both natural kind and artifact categories.

In order to explore the various possible reasons why some items should not follow a clean threshold function but instead should be scattered above and below the function, a regression function was fitted to the data from all 17 categories, (one category was excluded for technical reasons), and the residual categorization probability was calculated for each item. The items with categorization probability significantly higher or lower than that expected for their typicality were examined in more detail, and a number of hypotheses suggested themselves to account for the variation. First, there were a number of very unfamiliar items such as Euglena, or Lamprey, which had categorization probability higher than expected from Typicality. Typicality ratings are known to be affected by familiarity (Barsalou, 1985; Hampton & Gardiner, 1983). It is therefore quite likely that low familiarity with an item may depress its Typicality without affecting its categorization.

On the other hand there were items with lower categorization probability than expected, which appeared to be semantically associated with the category, but not actually category members. Examples were Orange Juice as a Fruit, or Egg as an Animal. Bassok and Medin (1997) have shown that semantic associatedness can give a sense of similarity, and it is not unreasonable to suppose that Typicality ratings may also reflect associatedness to an extent that is not seen in categorization itself.

Two further hypotheses were related to the distinction that Rips, Keil and others have stressed — namely the distinction between the surface appearance of objects, and their deeper nature. Some items bear a superficial resemblance to a category to which they do not belong — a whale as a Fish is perhaps the best known example. Other items bear little resemblance to the category to which they *do* belong — as might be the case for tomatoes and Fruit. It may be expected that items that are *technically not members* should have lower category probability than expected, while those with are *only technically members* should have higher probability than expected.

A final hypothesis concerned the effect of contrast categories on typicality and categorization. Similarity to a prototype may be calculated without regard to any contrasting or overlapping categories of which the item may be a member. Categorization however may proceed in a more contrastive manner, in that people may prefer to categorize each item in just one category (as in the *mutual exclusivity principle*, adopted by young children in word learning — Clark, 1973). If an item is a better member of some contrasting or overlapping category, then perhaps its categorization probability would be less than expected from its typicality.

These various hypotheses were collected together and tested in a rating questionnaire which was administered to twenty students at the University of Chicago. From this questionnaire, variables were computed for each item, corresponding to its Unfamiliarity, the degree to which it was Only Technically a member, or Technically Not a member, the degree to which it was judged a Part or Property rather than a true member, and the degree to which it also belonged in a Contrast category. These five new variables were entered into a regression to predict residual categorization probability when the effect of Typicality had been removed. Four of the five variables proved to be significant predictors, in the expected direction. Items that were Unfamiliar, or were Only Technically members, were associated with positive residuals — they were more likely to be categorized positively than warranted by their typicality. Items that were associated parts or properties, or that were Technically Not members were associated with negative residuals — they were less likely to be categorized positively than was warranted by their typicality. The Contrast variable had no overall predictive effect on residual categorization probability.

A subsequent analysis compared the 4 biological categories (Animal, Bird, Fish and Insect), with the 5 artifact categories (Clothing, Furniture, Kitchen Utensil, Ship and Vehicle). It was found that the two "Technical" predictors were significant for the biological categories, but not for the artifacts. On the other hand, the Contrast category predictor was significant only for the artifact categories. This difference is consistent with the fact that people may be influenced by biological classification in the zoological categories, but that no corresponding theory exists for artifacts. Similarly, artifacts often fall into overlapping categories (a knife may be either a tool, a weapon or a kitchen utensil), whereas biological categories are usually mutually exclusive. Hampton (1997) concluded that there were few systematic deviations from monotonicity and many of them could be accounted for by the effects of familiarity or associatedness on typicality ratings. There was also evidence that typicality gives less weight to "technical" or deeper aspects of objects than does categorization.

WHAT ROLE DOES SIMILARITY PLAY?

In this paper I have suggested that similarity-based categorization is in fact a widespread phenomenon, affecting not only the common everyday use of categories, but also people's reasoning processes about those categories. It would be foolish to argue that all of our categories are constructed on the basis of putting similar things together. We would certainly have made little progress culturally or scientifically if our conceptual repertoire were limited to such categories. How then can the evidence for similarity-based categorization be squared with this notion that our concepts should *not* be based on similarity?

There are two issues here to be kept separate. The first is that the world contains important distinctions that are not always immediately obvious in the outward appearance of objects. Two mushrooms may be very similar, but whereas one makes a tasty meal, the other is deadly poisonous. A crude view of similarity-based categorization would argue that we could never learn this distinction, since it would require forming a category that cuts across the way things appear to us perceptually. This view is to take *perceptual* (in fact usually *visual*) similarity as the only meaningful way of defining similarity. Perceptual similarity is indeed a very powerful and salient factor in our thinking, and it probably represents the "prototypical" or default way in which we understand similarity.

(It was not so much the principle of *similarity-based* categorization that Rips (1989) was attacking, so much as the notion of categorization based on *resemblance in appearance*.)

There is however a more powerful way to treat similarity, in which any dimension may enter into the computation of similarity. We might then talk of "deep similarity" as opposed to "surface similarity". If some subtle morphological characteristic of the mushrooms provided a clear predictor of the effects of eating them, then this characteristic would be given a very high weight in the computation of similarity for the purpose of culinary classification. After all there is very little similarity in the effects of eating the two mushrooms, and this factor would be sufficiently important to carry great weight in determining categorization.

The first point is therefore that similarity must be broadened to encompass a range of semantic information that goes well beyond the perceptual appearance of objects. When this is properly understood, it is clear for example why whales should not be fish. When examined more closely, when their behavior is observed and their internal organs (lungs, warm blood, brains etc.) are inspected, their similarity to other mammals, and dissimilarity from fish becomes quite obvious. There is no need for a theory of evolution to make this observation, just a curiosity about the way things are.

The second point is that over and above the ability to use similarity as the basis for categorization, we have the capacity to think in a more precise logical fashion. We can define explicit terms such as Prime Number or Triangle, or we can define explicit goals to be satisfied (as in Barsalou's ad hoc categories). If told a categorization rule, we can readily apply it to the world, and indeed there is a growing body of results which suggests that if asked to *invent* a categorization scheme we have a strong bias for rules based on single dimensions (Medin, Wattenmaker & Hampson, 1987).

This type of more axiomatic thought has obviously led to the huge success of mathematics and the mathematical sciences, and by its nature it makes little use of similarity. Scientific concepts tend to form all-or-none categories, which can enter into logical relations and scientific laws with absolute certainty. Before the days of numerical taxonomy, it was considered an essential requirement for classification schemes that they should be based on monothetic criteria. Debate centred on which were the most appropriate dimensions or features with which to create subclasses, and the value of a classification was to be found in the theoretically interesting generalisations that it permitted one to make.

What should be obvious to most psychologists who have attempted to study this more "advanced" type of thought is that it is actually very *difficult* for most people. School teachers have to spend hours and hours of patient explanation to get the majority of students to understand the principles of mathematics or scientific laws and their concepts, and the majority of the population never succeed in mastering the necessary skills in more than a rudimentary form. From the earliest days of experimental psychology it has been shown that people are poor at following the abstract logic of syllogisms, conditionals, or probability. They are also poor at using analogy in problem solving unless surface similarity helps to cue the appropriate connection. Arguments that similarity-based categorization is inadequate since it cannot form a solid foundation of concepts for logic and reasoning are therefore founded on a dubious premise — namely that most people have such a foundation readily available to them. It is perhaps more realistic to suppose that similarity forms the basis of most people's concepts most of the time, and that some individuals, with a lot of training and with the advantage of the cultural transmission of ideas from great thinkers of the past are able to develop more advanced thinking skills in particular domains. Dimly remembered lessons may lead us to believe that our concepts are clearer than they really are — or to defer to experts as keepers of the truth. However for everyday purposes we are content to continue putting together things that are (superficially or deeply) similar. After all, such a system serves us perfectly well for most daily purposes.

A similar point can be made about those concepts that reflect deeper theoretical information, such as many biological or natural kind terms. Through the evolution of our culture and its interest in scientific knowledge, we have (as a culture) developed sophisticated concepts such as mammal, vertebrate or insect, and the proper definition of these terms requires educated attention to scientifically relevant dimensions of the creatures in question, and may often fly in the face of superficial resemblance in the appearance of objects. As responsible members of our linguistic and cultural communities we feel bound to defer to experts in the correct application of these terms, at least in discourse contexts where "correct" classification matters. This "linguistic division of labour" has been noted among others by Putnam (1975). Medin and Ortony (1989) describe the same situation using the notion of "psychological essentialism" — the common belief that many natural kinds have an essence by which they can be correctly classified, even though that essence and how to detect it may be unknown to the lay person. My point is that although we may defer to experts and correct definitions when the context requires, we are also very willing to fall back on a similarity-based concept of many natural kind terms for other purposes. Studies of natural kinds (e.g. Hampton, 1995; Kalish, 1995; Malt, 1994) have shown that people are equally happy to think of natural kind categories as showing family resemblance structure, and of categorization in such categories as allowing for degrees of membership depending on similarity to known typical examples.

ACKNOWLEDGEMENTS

The author acknowledges support from the British Academy, the Nuffield Foundation, the University of Chicago and French Ministry for Higher Education and Science. Thanks also to Lawrence Barsalou, Daniele Dubois, Zachary Estes, John Gardiner, Dedre Gentner, Douglas Medin, Gregory Murphy, Lance Rips, Brian Ross and Stephen Sloman. An earlier version of this paper was presented at SimCat 1997, Interdisciplinary Workshop on Similarity and Categorisation in Edinburgh, November 1997.

REFERENCES

Ashby, F.G., & Gott, R.E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 33-53.

Barsalou, L.W. (1983). Ad hoc categories. *Memory and Cognition, 11*, 211-227.

Barsalou, L.W. (1985). Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded Structure in Categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Bassok, M., & Medin, D.L. (1997). Birds of a feather flock together: Similarity judgments with semantically rich stimuli. *Journal of Memory and Language, 36*, 311-336.

Clark, E.V. (1973). Meanings and Concepts. In J.H.Flavell, & E.M.Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (pp 787-840). New York: Wiley.

Hampton, J.A. (1982). A Demonstration of Intransitivity in Natural Categories. *Cognition, 12*, 151-164.

Hampton, J.A. (1988). Overextension of conjunctive concepts: Evidence for a Unitary Model of Concept Typicality and Class Inclusion. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 12-32.

Hampton, J.A. (1995a). Similarity-based categorization: the development of prototype theory. *Psychological Belgica*, 35, 103-125.

Hampton, J.A. (1995b). Testing Prototype Theory of Concepts. *Journal of Memory and Language, 34*, 686-708.

Hampton, J.A. (1996a). Conceptual Combination: Conjunction and Negation of Natural Concepts. *Memory and Cognition.*

Hampton, J.A. (1996b) The relation between categorization and typicality: an analysis of McCloskey & Glucksberg's (1978) data. Unpublished report.

Hampton, J.A. (1997) Similarity-based Categorization and the Fuzziness of Natural Categories. *Under review.*

Hampton, J.A., & Gardiner, M.M. (1983). Measures of Internal Category Structure: a correlational analysis of normative data. *British Journal of Psychology, 74,* 491-516.

Kalish, C.W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory and Cognition,* 23, 335-353.

Keil, F.C. (1989). *Concepts, Kinds, and Cognitive Development,* Cambridge, MA: MIT Press.

Keil, F.C., & Batterman, N. (1984). A Characteristic-to-Defining Shift in the Development of Word Meaning. *Journal of Verbal Learning and Verbal Behavior,* 23, 221-236.

Malt, B.C. (1994). Water is not H_2O. *Cognitive Psychology,* 27, 41-70.

McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory and Cognition, 6,* 462-472.

Medin, D.L., & Ortony, A. (1989). Psychological essentialism. In S.Vosniadou & A.Ortony (Eds.), *Similarity and analogical Reasoning* (pp. 179-195). Cambridge: Cambridge University Press.

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

Medin, D.L., Wattenmaker, W.D., & Hampson, S.E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology,* 19, 242-279.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Nosofsky, R.M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14,* 700-708.

Putnam, H. (1975). The meaning of 'meaning'. In Mind, language and reality, volume 2: Philosophical papers. Cambridge: Cambridge University Press.

Rips, L.J. (1989). Similarity, typicality and categorization. In S.Vosniadou & A.Ortony (Eds.), *Similarity and Analogical Reasoning.* Cambridge: Cambridge University Press.

Rips, L.J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General, 122,* 468-486.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192-232.

Rosch, E., & Mervis, C.B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and process in semantic Memory: A featural model for semantic decisions. *Psychological Review, 81,* 214-241.

Sloman, S. (1997). Feature inheritance in inductive reasoning. *Manuscript under review.*

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293-315.

van Mechelen, I., Hampton, J.A., Michalski, R.S., & Theuns, P. (Eds.) (1993). *Categories and Concepts: Theoretical Views and Inductive Data Analysis.* London: Academic Press.

Zadeh, L. (1965). Fuzzy sets. *Information and control, 8,* 338-353.

WHY CONCEPTUAL COMBINATION IS SELDOM ANALOGY

Mark T. Keane

Dept. of Computer Science, University College Dublin,Dublin 4, Ireland.

Fintan J. Costello

Dept. of Computer Science, Trinity College Dublin,Dublin 2, Ireland.

ABSTRACT

Structure-mapping is fast emerging as a unifying principle for a variety of different phenomenon; including analogy, metaphor, similarity and conceptual combination. In this paper, we argue that it is inappropriate to extend this idea to conceptual combination, as has been done in the dual-process theory (see Wisniewski, 1997a, 1997b). There are theoretical and empirical grounds for taking up this position. We propose an alternative account based on the constraint theory of combination, which sees the interpretation of concept combinations as one of satisfying multiple constraints of diagnosticity, plausibility and informativeness. This theory, which we would like to advertise as being the truth, does not use structure-mapping.

INTRODUCTION

Structure-mapping or structural alignment is fast emerging as a unifying principle for a variety of different phenomena: including analogy (e.g., Gentner, 1983; Holyoak & Thagard, 1995; Keane, 1988; Keane, Ledgeway & Duff, 1994), metaphor (e.g., Gentner, 1982; Gentner & Wolff, in press; Veale & Keane, 1994, 1997), similarity (e.g., Markman & Gentner, 1993a, 1993b; Markman & Wisniewski, 1997; Goldstone, 1994; Goldstone & Medin, 1994) and conceptual combination (Wisniewski, 1996, 1997a, 1997b; Wisniewski & Markman, 1993). In this paper, we argue that it is inappropriate to extend this structure-mapping account to conceptual combination; that is, to the process that enables people to interpret novel combinations like *horse bird, river chair* and so on.

Structural alignment is a process that matches the relational structure of two domains of knowledge (e.g., concepts or stories) in accordance with the systematicity principle (Gentner, 1983). This idea has been instantiated quite precisely in a number of computational models including the Structure Mapping Engine (Falkenhainer, Forbus & Gentner, 1986, 1989; Forbus & Oblinger, 1990; Forbus, Ferguson & Gentner, 1994), the Incremental Analogy Machine (Keane, 1990, 1997; Keane & Brayshaw, 1988; Keane et al., 1994), ACME (Holyoak & Thagard, 1989) and LISA (Hummel & Holyoak, 1997). In the context of conceptual combination, structural alignment is used to explain the generation of certain classes of interpretation that are produced to novel compounds.

In the remainder of this paper we argue against this proposal[1]. First, we describe the dual-process theory in some detail. Second, we object to this account with an alternative account called the constraint theory. Third, we outline evidence favouring the constraint theory over dual-process theory.

DUAL-PROCESS THEORY

Dual-process theory (Wisniewski, 1997a, 1997b) proposes that two main mechanisms underlie conceptual combination: structural alignment and scenario formation. Each of these processes is responsible for explaining the different types of interpretation that people produce. Structural alignment is proposed to explain property interpretations where an property from one concept is asserted of the other (e.g., an *elephant fish* is a big fish). It also accounts for hybrids where the interpretation is some combination of the properties of both concepts; e.g., a drill screwdriver is two-in-one tool with features of both a drill and a screwdriver. Scenario formation is very like Murphy's (1988; Cohen & Murphy, 1984) concept-specialisation mechanism and is used to explain relational interpretations (e.g., a night flight is a flight taken at night). We will concentrate on the structural alignment mechanism here as it is our main concern.

---

[1] There are circumstances under which analogy is certainly used to interpret combinations; for instance, it is hard to explain how Irangate could be interpreted without using Watergate by analogy (see Shoben, 1989). However, this type of interpretation is uncommon and cannot account for most of the interpretations normally produced.

The structural alignment process is similar to analogical structure-mapping (Gentner, 1983; see Keane, 1993, for a review). To interpret a given compound phrase the structural alignment process compares the two constituent concepts, and on the basis of that comparison selects an alignable difference to transfer from one conceptto the other. When two concepts are compared a number of different relationships can be found between their parts: commonalities (where both slot and value match), alignable differences (where both slot and value match) and non-alignable differences (where both slot and value match; see Figure 1). It is the values that are found in alignable differences that are used in property interpretations; for example, "an elephant fish is a big fish" is produced by comparing the concepts "elephant" and "fish", noticing that the "elephant "and "fish" share the dimension SIZE but have different values on that dimension, and transferring the alignable difference BIG from "elephant" to "fish". When a single aligned property is selected for transfer, a property interpretation is produced; if multiple properties are transferred, a hybrid interpretation results. The diagnosticity of a property may have a role in choosing between competing alignable differences, if more than one is available (Wisniewski, 1997a). One important prediction made from this alignment mechanism is that property interpretations should increase in frequency when the constituent concepts of a combination are similar (and hence, easy to align). This prediction has been confirmed in several studies (Wisniewski & Markman, 1993; Markman & Wisniewski, 1997). Dual-process theory is a well-developed account that makes several novel and interesting predictions about conceptual combinations, many of which have been confirmed empirically. However, we believe that structural alignment is not used in conceptual combination but have, in its stead advanced a theory, that can generate property, hybrid and relational interpretations using a very different set of mechanisms guided by certain high-level constraints. In the next section, we briefly describe this theory before describing some evidence that supports it but does not favour structural alignment.

	Elephant		Fish	
Commonalities	class	: living thing	class	: living thing
Alignable differences	size	: big	size	: small
	colour	: grey	colour	: silver
Non-alignable differences		: has trunk		
				: has fins

Figure 1. The Different Relationships that Occur When Two Concepts Are Aligned

THE CONSTRAINT THEORY

Constraint theory (Costello, 1996; Costello & Keane, 1997a, 1997b, 1998) describes conceptual combination as a process which constructs representations that satisfy three constraints of diagnosticity, plausibility and informativeness. These constraints derive from the pragmatics of compound interpretation and use (Grice, 1975; see Costello & Keane, 1997b, for details). In this section we describe the three constraints which the theory proposes; the specific algorithm for building representations that satisfy these constraints is described elsewhere (see Costello, 1996; Costello & Keane, 1997b).

The *diagnosticity constraint* requires the construction of an interpretation containing diagnostic properties from each of the concepts being combined. The diagnostic properties of a concept are those which occur often in instances of that concept and rarely in instances of other concepts (similar to Rosch's, 1978, cue validity). Diagnosticity predicts that the interpretation "a cactus fish is a prickly fish" is preferable to "a cactus fish is a green fish" because PRICKLY is more diagnostic of cactus than GREEN. Diagnosticity also identifies the focal concept or central concept which an interpretation is about; the focal concept of an interpretation is defined to be that part of the interpretation which possesses the diagnostic properties of the head noun of the phrase being interpreted.

The *plausibility constraint* requires the construction of an interpretation containing semantic elements which are already known to co-occur on the basis of past experience. The plausibility constraint ensures that interpretations describe an object (or collection of objects) which could plausibly exist. Plausibility would predict that the interpretation "an *angel pig* is a pig with wings on its torso" would be preferable to "an *angel pig* is a pig with wings on its tail", because prior experience suggests that

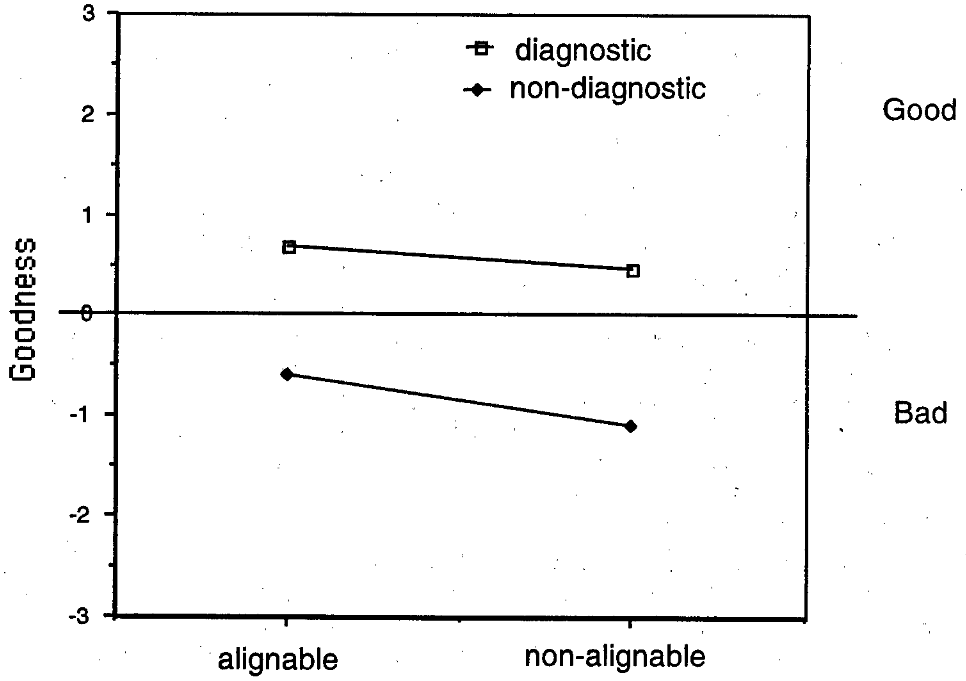


Figure 2. Mean Goodness Ratings for Different property Interpretations from Costello & Keane (1998).

wings are typically attached to the centre of gravity of an object (see also Downing, 1977).

The *informativeness constraint* requires the construction of an interpretation which conveys a requisite amount of new information. Informativeness excludes feasible interpretations that do not communicate anything new relative to either constituent concept; for example, "a *pencil bed* is a bed made of wood" is a feasible interpretation for "pencil bed" but no one presented with this compound has ever produced it as an interpretation (see Costello & Keane, 1997a). Together these three constraints account for the range of different combination types that have been observed: each combination type represents a different way of satisfying the constraints.

Empirical support for constraint theory comes from analyses of the rates of different interpretation-types produced to combinations involving constituents of different classes (e.g., artifacts, natural kinds, superordinates and basic-level concepts; see Costello & Keane, 1997a, 1997b). However, the theory also makes a novel prediction on the frequency of property interpretations in so-called called reversed-focal interpretations. In *reversed-focal* interpretations, the focal concept is the modifier concept (i.e., the first word) rather than the head (i.e., the second word); for instance, "a *chair ladder* is a chair that is by necessity used as a ladder" (see also Gerrig & Murphy, 1992; Wisniewski & Gentner, 1991). In constraint theory, the referent of an interpretation is identified by the diagnostic properties of the head concept of the phrase interpreted. Therefore, the theory predicts that reversed-focal interpretations should involve the diagnostic properties of the head being mapped to the modifier. In short, that reversed-referent interpretations will be property interpretations. Costello & Keane (1997a) have found that this is indeed the case, that while property interpretations were in general less frequent (around 30%) than relational interpretations (around 50%), for reversed-focal interpretations this pattern was reversed: around 50% of reversed-focals were property-mappings, with around 30% relational interpretations.

The constraint theory has also been implemented in a running computational model that has been tested on a large number of combinations; the C^3 model (Constraints on Conceptual Combination).

EVIDENCE AGAINST ALIGNMENT

Both the constraint theory and alignment theory speak to a common corpus of empirical evidence and each make their own predictions about certain novel phenomena. The difficulty is in finding evidence that decides between the two theories.

We know of only one piece of evidence that appears to present some difficulties for the predictions of the dual-process theory; namely, Costello & Keane's (1998) study of people's judgement of property interpretations involving properties that were systematically varied in terms of their alignability and diagnosticity. This experiment made use of a goodness judgement task for a set of property interpretations to noun-noun compounds. In the main experiment, participants were given four different property interpretations of a novel combination, each of which reflected one of the logical possibilities involving the two variables. For the novel combination "bumblebee moth", for example, participants received the following four possible interpretations:

Bumblebee moths are
(a) moths that are black and yellow
(aligned diagnostic)
(b) moths that are the size of a bumblebee
(aligned non-diagnostic)
(c) moths that sting
(non-aligned diagnostic)
(d) moths that fertilise plants
(non-aligned non-diagnostic)

Participants then rated the goodness of these meanings for the combination, using a seven-point scale (from -3 to +3). The interpretations used in this main experiment were constructed based on analyses from two pre-test experiments. In Pre-test 1, alignable and non-alignable differences for the concepts in each noun-noun phrase were gathered (using

Markman & Wisniewski's, 1997, methodology). In Pre-test 2, the diagnosticity of these selected alignable and non-alignable properties were determined in a rating study.

The results showed that people prefer property interpretations using non-alignable properties (if they are diagnostic) to alignable differences (if they are not diagnostic; see Figure 2). Notably, this experiment clearly shows that diagnostic, non-alignable properties support good property interpretations. This alignment account predicts that alignable properties will always be preferred.

CONCLUSIONS

In this paper, we have argued that structure mapping does not be extend to an account of conceptual combination but that other mechanisms provide a better account. It could be argued that certain parts of constraint theory might be handled by an alignment mechanism (e.g., the plausibility constraint is a likely candidate). We would resist such a proposal, if only to clarify the different sides in the debate. But, there are broader reasons for preferring a pure constraint account. That is, it seems to us that the pragmatics of understanding conceptual combinations are quite different to that which hold in analogy, and that, as such, there should be no reasonable expectation for analogical processes to play a role.

REFERENCES

Cohen, B. & Murphy, G. L. (1984). Models of concepts. *Cognitive Science, 8*, 27-58.

Costello, F.J. (1996). *Noun-noun conceptual combination: The polysemy of compound phrases.* Doctoral dissertation, submitted to the University of Dublin, Trinity College, Ireland.

Costello, F.J., & Keane, M. T., (1997a). Polysemy in conceptual combination: Testing the constraint theory of combination. In *Nineteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Costello, F.J., & Keane, M. T., (1997b). *Efficient creativity: Constraints on conceptual combination.* Submitted for publication.

Costello, F.J., & Keane, M. T., (1998). *Alignment Versus Diagnosticity in Conceptual Combination.* Seventh Irish Conference on AI and Cognitive Science. University College, Dublin, Ireland.

Downing, P. (1977). On the creation and use of English compound nouns. *Language, 53*(4), 810-842.

Falkenhainer, B., Forbus, K.D., & Gentner, D. (1986). Structure-mapping engine. *Proceedings of the Annual Conference of the American Association for Artificial Intelligence.* Washington, DC: AAAI.

Falkenhainer, B., Forbus, K.D., & Gentner, D. (1989). Structure-mapping engine. *Artificial Intelligence, 41*, 1-63.

Forbus, K.D., Ferguson, R.W., & Gentner, D. (1994). Incremental structure mapping. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Forbus, K.D., & Oblinger, D. (1990). Making SME greedy and pragmatic. *Twelfth Annual Conference of the Cognitive Science Society.* Hillsdale: Erlbaum.

Gentner, D. (1982). The role of analogy in science. In D.S. Miall (Ed.), *Metaphor: Problems and perspectives.* Sussex: Harvester.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Gentner, D. & Wolff, P. (1997). Metaphor and knowledge change. In A. Kasher & Y. Shen (Eds.), *Cognitive aspects of metaphor.* Amsterdam: North Holland.

Gerrig, R. J. & Murphy, G. L. (1992). Contextual influences on the comprehension of complex concepts. *Language and Cognitive Processes, 7*(3-4), 205-230.

Goldstone, R.L. (1994). Similarity, interactive activation and mapping. *Journal of Experimental Psychology: Language, Memory & Cognition, 20*, 3-28.

Goldstone, R.L. & Medin, D.L. (1994). Time course of comparison. *Journal of Exper-*

imental Psychology: Language, Memory & Cognition, 20, 29-50.

Grice, H.P. (1975). Logic and conversation. In P. Cole and J.L. Morgan (Eds.), *Syntax and semantics (vol 3): Speech acts.* New York: Academic Press.

Holyoak, K.J., & Thagard, P.R. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295-355.

Holyoak, K.J., & Thagard, P.R. (1995). *Mental Leaps.* Cambridge, MASS: MIT Press.

Hummel, J. E. & Holyoak K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review.*

Keane, M.T. (1988). *Analogical problem solving.* Chichester: Ellis Horwood (Simon & Schuster in N.America).

Keane, M.T. (1990). Incremental analogising: Theory and model. In K.J. Gilhooly, M.T. Keane, R. Logie & G. Erdos (Eds), *Lines of thinking: Reflections on the psychology of thought.* Vol. 1. Chichester: John Wiley.

Keane, M.T. (1993). The cognitive processes underlying complex analogies: Theoretical and empirical advances. *Ricerche di Psicologia, 17,* 9-36.

Keane, M.T. (1997). What makes an analogy difficult ?: The effects of order and causal structure in analogical mapping. *Journal of Experimental Psychology: Language, Memory & Cognition, 23,* 946-967.

Keane, M.T., & Brayshaw, M. (1988). The Incremental Analogy Machine: A computational model of analogy. In D. Sleeman (Ed.), *Third european working session on learning.* London: Pitman/San Mateo, Calif.: Morgan Kaufmann.

Keane, M.T., Ledgeway, T, & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science, 18,* 287 - 334.

Markman, A. B., & Gentner, D. (1993a). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32*(4), 517-535.

Markman, A. B., & Gentner, D. (1993b). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*(4), 431-467.

Markman, A.B., & Wisniewski, E. J. (1997). Same and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Language, Memory & Cognition, 23,* 54-70.

Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science, 12*(4), 529-562.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.) *Cognition and categorization.* Hillsdale, NJ: Erlbaum.

Shoben, E. J. (1993). Non-predicating conceptual combinations. *The Psychology of Learning and Motivation, 29,* 391-409.

Veale, T. and Keane, M. T. (1994). Belief modelling, intentionality and perlocution in metaphor comprehension. *Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum.

Veale, T. and Keane, M. T. (1997). The competence of sub-optimal structure mapping on "hard" analogies. *IJCAI'97:International Joint Conference on Artificial Intelligence.* Los Altos: Morgan Kaufmann.

Veale, T. and Keane, M. T. (1997). Principle Differences in Structure Mapping. *Analogy'98.* Sofia, Bulgaria

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language, 35*(3), 434-453.

Wisniewski, E. J. (1997a). Conceptual combination: Possibilities and esthetics. In T. B. Ward, S. M. Smith & J. Vaid (Eds.) *Creative thought.: An investigation of conceptual structures and processes.* Washington DC: American Psychological Association.

Wisniewski, E. J. (1997b). When concepts combine. *Psychonomic Bulletin & Review, 4*(2), 167-183.

Wisniewski, E. J. & Gentner, D. (1991). On the combinatorial semantics of noun pairs: Minor and major adjustments to meaning. In G. B. Simpson (Ed.) *Understanding word and sentence*. Amsterdam: North Holland.

Wisniewski, E. J. & Markman, A. B. (1993). The role of structural alignment in conceptual combination. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*. Boulder, CO.

ANALOGICAL PROBLEM-SOLVING BY CHIMPANZEES

David L. Oden,
La Salle University,
Philadelphia, PA. U. S. A.

Roger K. R. Thompson,
Franklin and Marshall College,
Lancaster, PA. U. S. A.

David Premack
Somis, CA. U. S. A.

ABSTRACT

Early research on the ability of chimpanzees to complete analogies provided evidence of that ability with regard to relationships between physical properties of geometric forms and with regard to functional relationships between common objects. Recent research, requiring not only completion of partially-constructed analogies, but also construction of an analogy from its elements, provided evidence for both abilities in the chimpanzee. However, the data suggest that the strategies used by the chimpanzee to solve such problems may be analogous, but not identical, to those used by humans.

Classical analogy problems involve perceptions and judgments about relations between relations Typically, the ability to solve such problems is regarded as a measure of computationally complex, reasoning at a developmentally sophisticated level (e. g., Goswami, 1991; Holyoak & Thagard, 1997; Piaget, 1977; Sternberg, 1977, 1982; Sternberg & Nigro, 1980; Vosniadou & Ortony, 1989). The question of whether such sophisticated reasoning is unique to humans has been a perennial topic for debate (cf., Darwin, 1871; Griffin, 1992; James, 1981/1890; Vauclair, 1996; Weiskrantz, 1985). There are techniques which allow one to systematic examine analogical reasoning and its component processes in species other than humans even though they lack the capacity for verbal report. For example, Gillan, Premack and Woodruff (1981) reported that a chimpanzee, Sarah, solved analogies which were instantiated using simple geometric forms presented in a 2 x 2 matrix format as shown in Figure 1. Here the stimuli A and A' exemplified a certain relation, (i.e., large vs. small), the stimuli B and B' exemplified the same relation but with different items (i.e., squares rather than circles), and "same" was the plastic token for this concept from the chimpanzee's artificial language (Premack, 1976). Thus, the array shown in figure 2 represents an analogy that a human might verbalize as, "large circle is to small circle as large square is to small square."

In one set of experiments, Gillan et al (1981) presented the chimpanzee Sarah with four items presented in the 2 x 2 format. If the arrangement constituted a true analogy then Sarah's task was to place her token for "same" in the center of the analogy matrix between the two arguments of that analogy. If the arrangement of items did not constitute an analogy then Sarah was correct if she placed her token for "different" in the center of the matrix. In another set of analogy problems Gillan et al (1981) presented Sarah with three terms of an analogy (i.e., A, A' and B) which were positioned according to the format described above. Sarah's

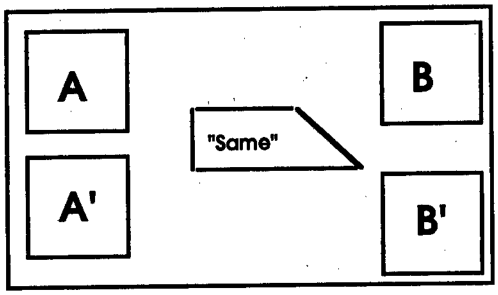


Figure 1. The 2 x 2 matrix format used by Gillan et al. (1981).

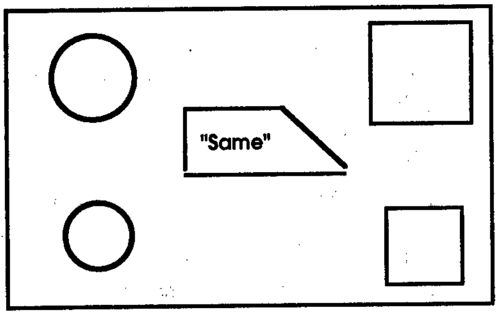


Figure 2. A geometric analogy in the 2 x 2 matrix format.

task was to select the appropriate fourth term (B') that was presented with another, but inappropriate, alternative.

In addition to solving these analogy problems involving arbitrary relations between geometric forms, Sarah also solved analogy problems (Gillan et al, 1981; Exp. 3) in which the common objects were used as the elements from which analogies could be constructed based on functional relations (e.g., padlock is to key as tin can is to can opener). In these functional analogy problems, the objects used to construct them were presented in the same matrix format as were the geometric problems. In both geometric and functional analogies, Sarah's task was essentially the same: To complete (or evaluate) a 2 X 2 arrangement of objects in which the relationship between the items in the left column was equivalent to the relationship between the items in the right column.

Gillan et al (1981) interpreted Sarah's successful performance on both geometric and functional analogy problems as reflecting her ability to reason about relations between relations. That is, she presumably established the relationship "same" (or "different") **between** the two sides of the analogy by first assessing and then comparing the relationships **within** each side. However, a close examination of the choices made by Sarah suggests that at least some of her apparently analogical based performances could have reflected far less sophisticated strategies.

Consider, for example, those problems which required Sarah to select a fourth item to complete a partially-constructed analogy. Sue Savage-Rumbaugh (personal communication, 1989) challenged the claim that Sarah employed true analogical reasoning to solve such problems. Specifically, Savage-Rumbaugh provided a detailed analysis of Sarah's performance which indicated that Sarah need not attend to the relationship instantiated by the A and A' elements on the left-hand side of the matrix. Savage-Rumbaugh showed how Sarah's choices could have been determined solely by a hierarchical set of featural matching rules by which she identified the choice item most like, if not identical to the single item (i.e., B) on the right-hand side of the matrix. Savage-Rumbaugh's analysis was compelling because it not only predicted the chimpanzee's correct choices, but also her errors. Furthermore, studies of analogical reasoning in 4- and 5-year old children (Alexander et al., 1989; Goswami, 1989) revealed that the less-proficient reasoners frequently resorted to such strategies.

Although Savage-Rumbaugh's featural similarity matching analysis has some heuristic value for explaining some of the Gillan et al (1981) results, it cannot account for Sarah's performance in other experiments in the same study which were designed explicitly to rule out physical matching or other associative processes for problem solving. Nevertheless, Savage-Rumbaugh's analysis is important because it raises fundamental questions regarding the conditions necessary for the expression of analogical reasoning abilities (cf., Oden, Thompson & Premack, 1990). For example, Sarah's analogical reasoning ability may only have been expressed in situations where it was mandated by the structure of the task. Consider, for example, the case of functional analogies. Faced with the question, "Padlock is to key as tin can is to...?" Sarah could not have chosen a can-opener instead of a paintbrush other than by comparing functional relationships .The utility of associative strategies in this task was precluded by the experimental design.

Recent advances in the study of analogies by a chimpanzee.

We present here a summary of extensive data analyses of more recent research conducted with Sarah on analogical problem solving tasks (Oden, Thompson & Premack, in preparation a; Oden, Thompson & Premack, in preparation b). These experiments were conducted in part to determine the boundary conditions for Sarah's analogical reasoning. For example, would Sarah use analogical reasoning spontaneously in situations where a simpler associative strategy would suffice? If so, then one could argue that she is predisposed, as are we humans, to reason about relations between relations; seeking out metaphor even when it is not explicitly required. Another goal of this research then was to determine whether Sarah could also construct, rather than merely complete, analogies. This task is substantially more demanding than those she faced in her earlier work. Completing or evaluating analogies requires one to compare relations which have been previously established; constructing analogies, however, requires one to seek out relations which reside among stimuli, but which have yet to be specified.

The materials used in this series of analogy tasks were similar to those used in the Gillan, et al (1981) geometric analogy problems. Sarah worked with an analogy board; a blue cardboard rectangle with an attached white cardboard cross, the arms of which extended across the length and width of the rectangle. This provided, at each corner of the rectangle, a recess into which stimuli could be placed to construct an analogy. Sarah's plastic token for the concept "same." was placed at the intersection of the display board's arms.

The experimental stimuli were squares of white cardboard, each with a geometric form stenciled on it. The forms varied in color (4), shape (3), size (2), and whether they were filled in with color or simply a colored outline. All possible combinations of these properties were used to create a pool of 48 different items which were used in the experiments reported here.

The following rules were used to select items for the analogies. A and A' differed with respect to a single dimension (size, color, shape or fill). B and B' also differed in this single dimension. A differed from B (and thus A' differed from B') on two dimensions, each different from the property distinguishing A and A'. For example, if A' represented a size transformation of A, then B might differ from A with respect to color and shape or shape and fill. Following these rules, a total of 612 unique combinations of 4 stimuli could be selected which, when appropriately placed on the board, would create an analogy. When experimental conditions required presentation of an additional (error) alternative choice item, this item (C) differed from B' along the dimension which was not used in constructing the analogy. For example, if the analogy was a "size x shape+fill", then C differed from B' in color.

Sarah worked with these materials under four conditions. In two conditions, she was required to complete partially-constructed analogies which were presented on the analogy board. In two other conditions, she was presented with an empty analogy board along with the appropriate stimulus items and had to construct an analogy from scratch. Throughout the study, a unique set of 4 analogy items was used on each trial.

General test procedures. A standard test procedure was used in all conditions. On each trial of a test session, the trainer placed the analogy board just inside the wire mesh of Sarah's home cage enclosure. The board contained either a partially-constructed analogy (Completion Conditions 1 & 2) or no stimuli at all (Construction Conditions 3 & 4). The stimuli which served as 'answer' alternatives were contained in a covered cardboard box which the trainer placed in front of the analogy board. After presenting the materials, the trainer left the room and recorded Sarah's behavior via a one-way mirror. Sarah's task was to open the alternatives box, make her selections and place the items in the empty recesses of the analogy board. Any unused items were either left in the box or, at Sarah's discretion, placed in a pie tin adjacent to the testing area. She then rang a small bell inside her enclosure, summoning the trainer back into the room.

In those sessions where the design called for differential feedback (Completion Condition 1), Sarah was praised and given a piece of fruit after each trial when she had completed an analogy. When she erred, she was mildly admonished and the trainer demonstrated the proper arrangement of stimuli but gave no food reward. In those sessions which called for non-differential feedback (Completion Condition 2; Construction Conditions 3 & 4), Sarah was praised and given a food reward for every trial regardless of her accuracy, unless she had left an unfilled space on her analogy board. In that case, the trainer pointed to the empty recess and instructed Sarah to "Do better next time." No other feedback was given on such trials. Under non-differential feedback, no particular problem-solving strategy is explicitly required, allowing the chimpanzee, if she is so inclined, to demonstrate spontaneous analogical reasoning (cf., Oden, Thompson & Premack, 1988).

DOES A CHIMPANZEE COMPLETE ANALOGY PROBLEMS ANALOGICALLY?

Condition 1: Completion with two alternatives. This condition was a replication of the forced-choice task used by Gillan et al. (1981), in which Sarah was required to select a single item (B') to complete a partially-constructed analogy. This condition was intended to familiarize Sarah with the new analogy board and stimulus items, and to provide a performance baseline. The analogy elements A, A' and B were placed in their appropriate positions on the board by the trainer. Two items, B' and an error alternative (C), were placed in the alternatives box. One session of twelve trials was run using differential feedback.

Three of the 12 trials could not be scored because one or more of the recesses on the analogy board were empty when the trainer was summoned by Sarah's bell. In two of these cases, this was the result of Sarah having dismantled the partially-constructed analogy to closely inspect the new stimulus materials. In the third case, both alternatives were laid on the floor beside the intact analogy board. Sarah succeeded in completing the analogy on 8 of the 9 trials which could be scored. This level of performance (89%; $p < .05$, Binomial test) compares favorably with the 75% overall accuracy reported in the original analogy studies (Gillan, et al., 1981).

Condition 2 : Completion with three alternatives. This condition was run to determine whether Sarah could not only select items necessary to complete an analogy, but also position them on the board so that the final product reflected an analogical arrangement. In this condition, the trainer placed only A and A' on the board. B, B' and C were placed in the alternatives box. Sarah's task was to select and properly arrange B and B' on the board. The arrangement of the items in the alternatives box was random. Four sessions of twelve trials each were run, using non-differential feedback.

Sarah completed an analogy on 22 of 48 trials (46%) , significantly more often than the 16% expected by chance. She selected the analogy pair (B, B') on 27 of 48 trials (56%; chance = 33%). On 22 of these 27 trials (81%; chance = 50 %) the selected items were placed on the board in the B/B' arrangement which completed the analogy begun with A/A'.

Sarah's overall success at completing analogies under this second condition, while statistically significant, was substantially lower than in Condition 1. Our examination of her relative success on the two components of this task (item selection and analogical placement) suggests that, for Sarah, the first component was the more difficult of the two. That is, although she selected the potential analogy choice pair on only 56% of the trials, once this pair was selected, Sarah arranged them analogically 81% of the time. Contrary to Savage-Rumbaugh's analysis of Gillan et als. (1981) initial results, the present data strongly suggest that Sarah's performance on analogy completion tasks was not significantly influenced by a simple matching strategy or other assessments of mere featural similarity. Rather, Sarah's performance was guided by the **relations** between features in the A/A' arrangement presented on her analogy board.

Her attention to relations is particularly striking given that non-differential reinforcement was used in Condition 2. This meant that she could have used any strategy whatsoever (including random selection and placement) to fill the analogy board. Nevertheless, she appears to have spontaneously adopted the relations between relations strategy. The next two conditions were intended to determine whether Sarah could detect and use relations to **construct** an analogy when presented with the necessary elements and an empty analogy board.

Will a chimpanzee construct analogies spontaneously?

Condition 3: Construction with four alternatives. In this condition, Sarah was presented with a completely empty analogy board and her alternatives box containing the four items necessary to construct an analogy. When Sarah placed the items in the recesses of her analogy board, non-differential reinforcement was given, regardless of whether their arrangement constituted a valid analogy. The criterion used for scoring her constructions was as follows. Sarah did not have to place the stimulus items originally designated by the investigators as A, A', B, B' in any particular recess. Any arrangement using these four elements was accepted as an analogy if A and B appeared together on one axis (row or column) of the board, and where A and A' appeared together on the alternative axis (column or row). This scoring rule was based on the property of an analogy that its elements and arguments may be interchanged in certain ways and still maintain analogical relations. For example, the construction "dog:cat::puppy:kitten" is as valid as "cat:kitten::dog:puppy" even though the relations expressed are rearranged. However, "cat:puppy::kitten:dog" would not be accepted as a valid analogy.

There were 24 possible arrangements of the items for a given trial, 8 of which (33%) would qualify as analogies. Sarah constructed valid analogies on 28 of 45 trials (62%), significantly more often than expected by chance. These results provide good evidence that Sarah constructed classical analogies using the same criteria as a human.

Did Sarah additionally understand the nature of the task before her? That is, did she intend to construct an analogy when she began a trial or did analogies unintentionally unfold as a necessary consequence of her initial choices? The answer to this question lays in the nature of her first two choices and their placement on the board. On approximately 90% of the trials, Sarah placed her first two choices in the same row or column on her analogy board, thereby determining whether an analogy could be completed.

With 4 alternatives, there were 12 possible ways that the first 2 items could be chosen. Eight of these combinations, when placed in the same row or column of the analogy board, constituted a "potential analogy" (i.e., they could become part of a valid analogy if the remaining items were arranged properly). Thus, Sarah could create, randomly, a potential analogy 67% of the time. But, in fact, her first two choices and placements produced potential analogies 82% (37/45 trials) of the time. Thus, we have evidence that Sarah exercised what might be called "foresight" in constructing her analogies. She essentially created the initial conditions that had been previously provided by the experimenters in Condition 2 of the completion task. On 76% (28/37) of these trials Sarah successfully completed the construction of a valid analogy. This level of success is consistent with her prior performances on the completion tasks reported here and by Gillan et al. (1981).

Condition 4: Construction with five alternatives. This condition was used to explore the effect of requiring an additional selection process as part of analogy construction. Recall that in Condition 3, the selection process proved to be more fragile than the arrangement process. We were curious whether Sarah, faced with this additional complexity, would resort to a simpler associative strategy or perhaps abandon all strategies in favor of random selection and placement. In this condition, Sarah was presented with an empty analogy board and her box of alternatives which contained four elements that could be used to construct an analogy, and a fifth, unusable item (C, the error alternative). As in Condition 3, Sarah's task was simply to fill the four empty spaces on the board for which she received non-differential feedback.

In this condition, Sarah constructed analogies on 21% of the trials. As expected, this level of performance was substantially lower than performance in the three preceding conditions, but it was nevertheless still significant ($p < .001$, binomial test). As before, we examined the sequence of Sarah's selections and placements to determine whether her performance truly reflected analogical reasoning or if it was the accidental byproduct of some simpler strategy. Two such strategies are considered below.

Strategy 1: Minimizing Featural differences. One possible strategy is that Sarah was guided by an appreciation of a more global pattern of relationships within an analogy, rather than the relations between particular pairs of items. We computed the total number of featural differences among members of the five 4-item sets which could be drawn from the larger 5-item set presented on each trial. According to the rules used to select those items, the subset which could be used to construct an analogy (A, A', B, B') would necessarily involve a minimum number of featural differences. However, another subset (C, A', B, B') also minimized the number of featural differences between its members

If Sarah were following a strategy of "minimize featural differences on the board," this would have led to completion of analogies in Condition 1. In Condition 2, this strategy would have led to the appropriate selection, but not necessarily to the appropriate arrangement, of items needed to construct an analogy. In the present condition, this strategy would have led to the selection of the potential analogy set. But it should also have led equally often to the selection of the set containing item C, the error alternative. In fact, Sarah selected the potential analogy set 46% of the time and selected the other "minimal-difference" set only 12% of the time (chance = 20%). Thus, Sarah was clearly not trying to simply maximize overall similarity among the four items placed on the board. It would be tempting, therefore, to conclude that the relationship between particular items (a prerequisite of analogical reasoning) was of significance to Sarah. However, an alternative strategy must be considered before accepting this conclusion.

Strategy 2: Exclusion of C, "the Odd man Out". It could be that Sarah adopted a strategy of **excluding** alternative C which possessed a single property (size, shape, color or fill) which was **not** shared with any of the other five-items. This strategy would have led Sarah to select the four items which could be used to construct an analogy, but only if they were arranged appropriately on the board. Given a selection of the appropriate items, one-third of their possible arrangements would meet our criteria, described previously, for an analogy. Using this one third proportion as an estimate of chance success, Sarah's performance was not statistically significant suggesting, therefore, that Sarah had not attended to relations between relations in this condition. However, a more detailed analysis of the temporal sequence in which Sarah placed the four items on the board led us to reject this pessimistic conclusion.

SARAH'S STRATEGY FOR CONSTRUCTING ANALOGIES.

Equating within-pair differences. As Sarah selected items and placed them on the board, she seems to have followed a strategy of equating the number of within-pair featural differences, independently of the physical nature of those differences. This strategy is illustrated in Figures 3a - 3d. Sarah consistently placed her first two choices on the same horizontal or vertical axis of the analogy board, as illustrated in Figure 3a. Here, B' (choice 2) and A (choice 1) have been placed respectively in the upper and lower recesses (i.e., a vertical axis) on the left hand side of the board. We can now describe Sarah's third and fourth choices as being placed adjacent to either her first or her second choices. In this example Sarah placed item C (choice 3) in the upper right-hand recess adjacent to her second choice (see Figure 3b). Sarah's fourth choice (A') was then placed in the lower right-hand recess adjacent to her first choice (see Figure 3c). Thus, Sarah's last two placements of her third and fourth choices could be described as creating two pairs as shown in Figure 3d. The number of featural differences within each pair is the same. That is, there is one featural difference in the B' & C pair created by Sarah's placements of her second and third choices. The A & A' pair created by her placements of her first and fourth choices similarly contains a single featural difference.

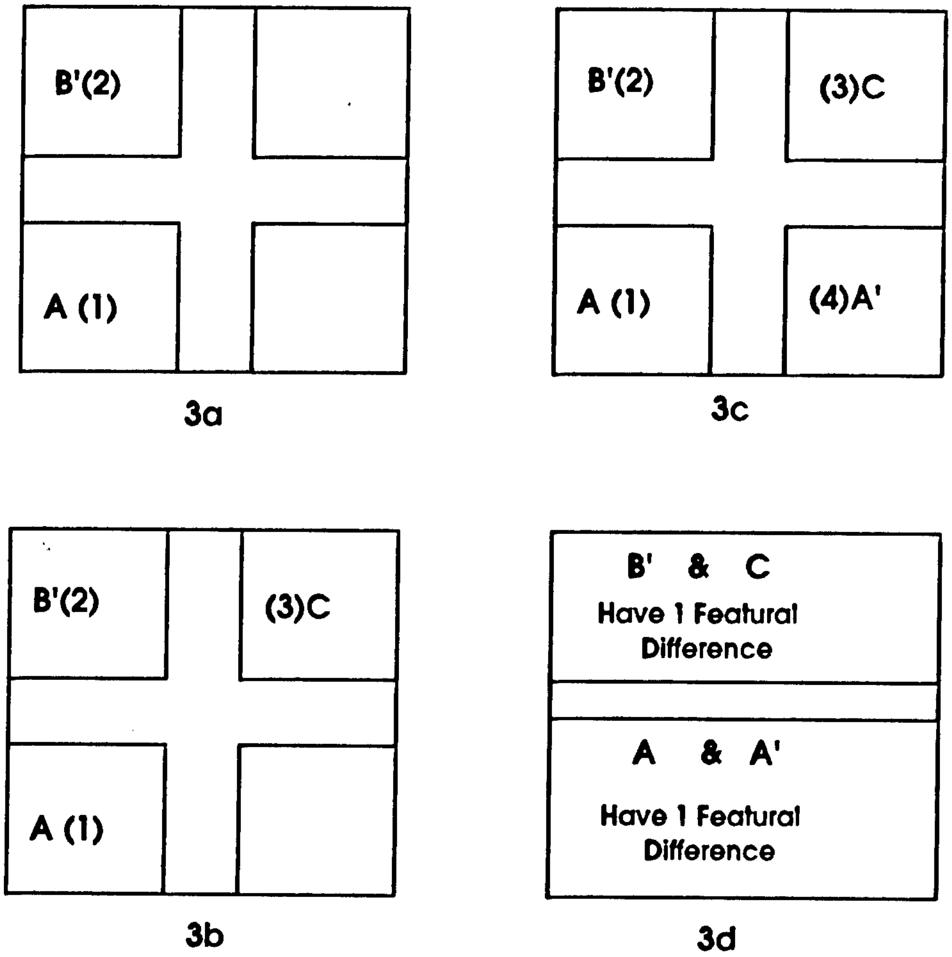


Figure 3. An illustrative sequence of Sarah's choices and placements in condition 4.

Each trial from Condition 4 of the analogy construction was analyzed in the manner described above (Oden et al., in preparation b). The expected frequencies of each combination of featural differences were obtained by determining the six possible outcomes given her two initial choices. The observed frequencies of pairings which equated within-pair differences significantly exceeded their expected frequencies.

Sarah apparently followed a strategy of numerically equating within-pair featural differences as she made her last two selections and placed them on the board. When Sarah placed her third choice next to one of the items already on the board the resulting number of within-pair featural differences tended to be subsequently matched within the pair created by her placing her fourth choice next to the remaining item.

We argue that this pattern of results reveals analogical reasoning; it involves reasoning about relations between relations. There is a difference, of course, between the strategy employed by Sarah and the a priori rules we used to construct analogies. Whereas we had attended to the nature of **specific** features, as well as their number, Sarah attended only to the number of featural differences. For example, we regarded a (color+shape) transformation as differing from a (size+fill) transformation. In Sarah's eyes these transformations were equivalent because they both entailed two featural differences. Thus, compared to our reasoning, Sarah's may lack rigor, but fundamentally, she still reasoned about relations between relations. We do not believe that Sarah's failure to attend to featural details beyond number reflects a fundamental constraint on her reasoning abilities. Recall that the results from Condition 2 of the completion task indicated that selection of items was a more difficult task than their arrangement. We believe that the decline in Sarah's performance in the present condition of the construction task resulted from the inherent complexity of the 5-item stimulus array with which she was presented.

SUMMARY

Collectively, the results from these four conditions not only confirm that an adult chimpanzee can solve analogies (Gillan et al., 1981), but also demonstrate that she does so spontaneously, even in situations where a simpler associative strategy would suffice.

In condition one we replicated Gillan et als. (1981) earlier findings which demonstrated that when faced with a partially constructed analogy problem Sarah, the same subject, successfully selected from two available choices that item which would complete the analogy. In condition 2 of the completion task, Sarah demonstrated conclusively that her performances was mediated by analogical relationships and not a simple associative similarity matching strategy. When presented with only two elements of a classical analogy problem she successfully chose from 3 alternatives the two elements necessary to complete the problem. More importantly however, was the finding that her spatial arrangement of these choices was guided by the relation initially established by the experimenters and not on the basis of mere similarity along any single physical dimension.

In conditions 3 and 4 we further demonstrated the same chimpanzee, Sarah, could not only complete, but also could construct analogies. When presented with a randomized grouping of elements from which an analogy could be constructed she proceeded to do spontaneously. When presented with the minimum of 4 elements she proceeded to arrange all of them in analogical fashion. When presented with 5 elements of which 4 could be used to construct an analogy she ignored the inappropriate item and successfully arranged the remaining items analogically. However, she did so in a manner analogous to, but not identical with that of her human experimenters. On the one hand, we had attended to both specific physical factors and their number in each within pair transformation. Sarah, on the other hand, attended to only the latter numerical dimension.

PRECURSORS FOR ANALOGICAL REASONING

Some investigators have argued that analogical reasoning is the common foundation (denominator) of much of human reasoning including logical inference (e.g., Halford, 1992). Our results confirm earlier reports (Gillan et al., 1981) that it is well within the capabilities of at least one adult chimpanzee. Might this capacity be expected in chimpanzees other than Sarah? Our answer is a qualified yes. Prior to her experience with formal analogical problem solving Sarah had mastered a conceptual matching task (Premack, 1978) which, at the age of 39 years, she still successfully performed under conditions of nondifferential reinforcement (Thompson, Oden & Boysen, 1998).

In the conceptual matching task a subject is required to match a pair of physically identical sample items (e.g., a pair of locks) with another pair of identical items (e.g., a pair of cups) as opposed to a pair on physically nonidentical items like, for example, a pencil and an eraser. Conversely, this latter nonidentical pair would be the correct match given another nonidentical sample pair such as a shoe and ball. Successful performance of the conceptual matching task described above involves the matching of relations between relations. It is then in essence a form of analogy in which all the arguments are provided for the subject. We believe, therefore, that any chimpanzee capable of performing the conceptual matching task possesses the computational cognitive foundations upon which formal analogical reasoning rests.

There is good evidence, however, that not all chimpanzees, can match relations between relations despite their success on physical matching tasks. Some prior experience with tokens which symbolize abstract same/different relations is apparently a necessary prerequisite for the explicit expression by a chimpanzee of their otherwise only implicit knowledge about relations between relations (Premack, 1983; Thompson et al. 1998). Presumably, experience with external symbol systems in some way provides the necessary representational scaffolding for the complex computational operations involved in solving problems involving conceptually abstract similarity judgments as in analogies (Clark & Thornton, 1997; Gentner & Markman, 1997; Sternberg & Nigro, 1980). Interestingly, this is, as yet, no evidence that old-world macaque monkeys can perceive, let alone judge, analogical relations (Thompson & Oden, 1996; Thompson & Oden, 1998; Washburn, Oden & Thompson, 1997).

CONCLUSION

The results described here on analogical problem solving by Sarah demonstrate that this chimpanzee is predisposed, as are adult humans, to reason about relations between relations. There was no evidence in the completion and construction analogy tasks summarized above that Sarah attempted to use a less efficient associative strategy, as can occur with young children (Alexander et al, 1989). If analogical reasoning is indeed a hallmark of human reasoning then its demonstration in a chimpanzee should not be surprising to anyone comfortable with a perspective on the origins of human cognition in which evolutionary and cultural factors are conjoined.

ACKNOWLEDGMENTS

We thank Bonnie Dennis and Teresa Anderson their assistance in data collection. The research reported here and preparation of this paper were supported by funds from the National Science Foundation (NSF-BNS 8418942 and NSF-IBN 9420756).

REFERENCES

Alexander, P. A., Willson, V. L., White, C. S., Fuqua, J. D., Clark, G. D., Wilson, A. F., & Kulikowich, J. M. (1989). Development of analogical reasoning in 4- and 5-year-old children. *Cognitive Development, 4*, 65-88.

Clark, A., Thornton, C. (1997). Trading Spaces: Computation, representation, and the limits of uniformed learning. *Behavioral and Brain Sciences, 20,* 57-90.

Darwin, C. (1871). The descent of man and selection in relation to sex. London, U.K.: Murray.

Gentner, D. & Markman, A. B. (1997). Structural mapping in analogy and similarity. *American Psychologist, 52,* 45-56.

Gillan, D. D., Premack, D., & Woodruff, G. (1981). Reasoning in the chimpanzee: I. Analogical reasoning. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 1-17.

Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cognitive Development, 4,* 251-268.

Goswami, U. (1991). Analogical reasoning: What develops? A review of research and theory. *Child Development, 62,* 1-22.

Griffin, D. R. (Ed.). (1992). *Animal thinking.* Chicago, IL.: Chicago University Press.

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development, 35,* 193-217.

Holyoak, K. J. & Thagard, P. (1997). The analogical mind. *American Psychologist, 52 (1),* 35-44.

James W. (1981/1890) *The principles of psychology, vol. I.* Cambridge, MA: Harvard University Press. (Original work published 1890).

Oden, D. L., Thompson, R. K. R., & Premack, D. (1988). Spontaneous transfer of matching by infant chimpanzees (*Pan troglodytes*). Journal of Experimental Psychology: Animal Behavior Processes.

Oden, D. L., Thompson, R. K. R., & Premack, D. (1990). Infant chimpanzees (*Pan troglodytes)* spontaneously perceive both concrete and abstract same/different relations. *Child Development, 61,* 621-631.

Oden, D. L., Thompson, R. K. R., & Premack, D. (In preparation a). A chimpanzee completes analogy problems analogically.

Oden, D. L., Thompson, R. K. R., & Premack, D. (In preparation b). Construction of analogies by a chimpanzee.

Piaget, J. (1977). *L'Abstraction reflechissante.* Paris, FR.: Presses Universitaires de France.

Premack, D. (1976). Intelligence in ape and man. Hillsdale, N. J.: Erlbaum Associates.

Premack, D. (1978). On the abstractness of human concepts: Why it would be difficult to talk to a pigeon. In S. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior.* Hillsdale, NJ: Erlbaum Associates.

Premack, D. (1983). The codes of man and beast. *The Behavioral and Brain Sciences, 6,* 125-137.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning.* Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1982). Reasoning, problem solving, and intelligence. In R. J. Sternberg, (ed.), *Handbook of human intelligence* (pp. 225-307). New York, N. Y.: Cambridge University Press.

Sternberg, R. J. & Nigro, G. (1980). Developmental patterns in the solution of verbal analogies, *Child Development, 51,* 27-38.

Thompson, R. K. R., & Oden, D. L. (1996). A profound disparity revisited: Perception and judgment of abstract identity relations by chimpanzees, human infants, and monkeys. *Behavioral Processes, 35,* 149-161.

Thompson & Oden, (1998). Why monkeys and pigeons, unlike certain apes, cannot reason analogically. *Proceedings of the Analogy '98 workshop on advances in analogy research: Integration of theory and data from the cognitive, computational, and neural Sciences.*

Thompson, R. K. R., Oden, D. L. & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal behavior processes, 23,* 31-43.

Vosniadou S. & Ortony, A. (Eds.). (1989). *Similarity and analogical reasoning,* Cambridge, England: Cambridge University Press.

Vauclair, J. (1996). *Animal cognition: An introduction to modern comparative psychology.* Cambridge, MA.: Harvard University Press.

Washburn, D. A., Thompson, R. K. R. & Oden D. L. (1997). *Monkeys trained with same/different symbols do not match relations.* Paper presented at the 38th Annual Meeting of the Psychonomic Society. Philadelphia, PA.

Weiskrantz, L. (Ed.). (1985). *Animal intelligence.* (Oxford Psychology Series No. 7). Oxford, U. K.: Clarendon Press.

ANALOGICAL REASONING IN CHILDREN

Usha Goswami,
Institute of Child Health, University College London, 30 Guilford St., London WC1N1EH, U.K.
E-mail: u.goswami@ich.ucl.ac.uk

Analogical thinking is the basis of much of our everyday problem solving. 'Analogy pervades all our thinking, our everyday speech and our trivial conclusions as well as artistic ways of expression and the highest scientific achievements' (Polya, 1957). The central role of analogy in human cognition underlines the importance of understanding the development of reasoning by analogy in children. However, until fairly recently, there was little interest in analogical development among researchers in child psychology.

This was because the most famous developmental psychologist, Piaget, had argued that analogical skills did not develop until early adolescence, and this conclusion had not been challenged. Rather than seeing analogy as a fundamental cognitive process, Piaget saw analogy as a sophisticated reasoning strategy that emerged after the primary years. The main reason was that, according to Piaget's general theory of logical development, the ability to see relations between relations (to use 'higher-order relations') was a hallmark of the final stage of logical reasoning, called the 'formal operational' stage. Formal operational reasoning required children to operate mentally on the results of simpler operations. A simpler operation was finding relations between objects (these simpler logical operations were called 'concrete operations'). As analogies required children to reason about relational similarity rather than about relations between objects, it appeared to be a typical formal operational skill.

Piaget's theory of logical development is the most widely-taught theory in cognitive developmental psychology and in education. It has also been used as a basis for research in many related areas (e.g., in theorising about the cognitive processes in reading development). If Piaget's conclusions about the relative mental sophistication of analogical reasoning turn out to be incorrect, then the implications for educational practice are immense.

Piaget's conclusions were based on experiments using a pictorial version of the standard test for analogical reasoning (used in IQ testing), the 'item analogy'. In item analogies, two items A and B are presented to the child, a third item C is presented, and the child is required to generate a D term that has the same relation to C as B has to A. Successful generation of a D term requires the use of the relational similarity constraint. For example, if the child is given the items '*cat is to kitten as dog is to* ?', she is expected to generate the solution term 'puppy'. The response 'bone', which is a strong associate of dog, would be an error. Another example is the analogy '*Bicycle is to handlebars as ship is to* ?'. Here the relation constraining the choice of a D term is 'steering mechanism', and so a child who offered the completion term 'bird' would not be credited with understanding the relational similarity constraint. Piaget's theory that analogical reasoning was absent in children until adolescence was based on item analogies such as these. Younger children tested by Piaget offered solutions like 'bird' to the *bicycle/ship* analogy, giving reasons like 'both birds and ships are found on the lake'. Piaget's interpretation of his research was that younger children solved analogies on the basis of associations. Children only became able to reason on the basis of relational similarity at around 11-12 years of age.

THE ROLE OF RELATIONAL FAMILIARITY IN ANALOGICAL DEVELOPMENT

Closer inspection of Piaget's experimental methods suggest a serious flaw, however. Piaget had not checked whether the younger children in his experiments understood the relations on which his analogies were based (relations such as 'steering mechanism'). Their failure to solve the item analogies in his experiments could thus have arisen from a lack of knowledge of the relations being used. Item analogies based on **unfamiliar** relations would obviously **underestimate** analogical ability.

The solution is to design analogies based on relations that are known to be highly familiar to younger children from cognitive developmental research. Simple **causal** relations such as *melting*, *wetting* and *cutting* are known to be understood between the ages of 3 and 4 years, and relations between real world objects such as *'trains go on tracks'* and *'birds live in nests'* are familiar to 4- and 5-year-olds. Item analogies such as *'playdoh is to cut playdoh as apple is to cut apple'* and *'bird is to nest as dog is to doghouse'* can thus be used to examine whether 3- to 5-year-olds have the ability to reason by analogy.

For this young age group, a picture-based version of the item analogy task was developed (Goswami & Brown, 1989, 1990). The task was presented as a 'game' about matching pictures. The children were shown a 'game board' with four slots for pictures, the slots being grouped in two pairs for the A:B and C:D parts of the analogy. As the children watched, the experimenter presented the first three terms of a given analogy (e.g., pictures of a bird [A], a nest [B], and a dog [C]). As the pictures were presented, the child was asked to name each one to ensure that they were familiar. The child was then asked to predict the picture that was needed to finish the pattern. This was intended to see whether children could generate an analogical solution spontaneously, without seeing the solution pictures.

Following this, the experimenter showed the child a choice of solution terms. For the *bird/dog* analogy, these were pictures of a *doghouse*, a *cat*, another *dog*, and a *bone*. The different choices were designed to test different theories of analogical development. The correct choice, which would indicate analogical ability, was the doghouse. The associative choice was the bone. Selection of the bone would be expected if younger children rely on associative reasoning to solve analogies, as Piaget had claimed. The other choices were a 'mere appearance match' choice (the second dog), and a semantic match (the cat). 'Mere appearance' matching is a term coined by Gentner (1989) to refer to the matching of object or 'surface' similarities when attempting to solve analogies (such as choosing another dog to match the dog in the C term). Gentner has suggested that younger children might rely on object similarity rather than relational similarity in reaching analogical solutions (Gentner, 1989).

The picture matching game showed that all children tested (4-, 5- and 9-year-olds) performed at levels significantly above chance in the analogy task, selecting the correct completion term 59%, 66% and 94% of the time respectively. There was no evidence of mere appearance matching. Although many younger children were shy of making predictions prior to seeing the solution choices, those who were more confident showed clear analogical ability on this measure as well. For example, when 4-year-old Lucas was given the analogy *bird is to nest as dog is to ?*, he first predicted that the correct solution was *puppy*. He argued, quite logically, "Bird lays eggs in her nest [the nest in the B-term picture contained three eggs] - dog - dogs lay babies, and the babies are - umm - and the name of the babies is puppy!" Lucas had used the relation *type of offspring* to solve the analogy, and was quite certain that he was correct. He continued "I don't have to look [at the solution pictures] — the name of the baby is puppy!" Once he looked at the different solution options, however, he decided that the *dog house* was the correct response.

The matching game also included a control task to ensure that the correct solution to the analogy was not simply the most attractive pictorial match for the C term picture. Here the children were simply shown the C term picture along with the correct solution term and the distractors, and were asked to choose which picture 'went best' with the C term picture. For example, the children were shown the picture of the dog, and were asked to choose the best match from the pictures of the doghouse, bone, second dog and cat. In this unconstrained task, the children were as likely to select the associative match (bone) as the analogy match (doghouse). Additionally, although the children readily agreed that another match could be correct in the control condition (9 year olds: 76%, 4 year olds: 82%), they were not so flexible in the analogy condition, where most of them said that only **one** answer could be correct (9 year olds: 89%, 4 year olds: 60%). This shows awareness of the relational similarity constraint that governs truly analogical responding. The children understood that the correct completion term for the analogy had to link the C and D terms by the **same** relation that linked the A and B terms. Notice that Lucas was using the relational similarity constraint when he generated the solution 'puppy' for the *bird/dog* analogy. This cognitive flexibility displays a full understanding of analogy, and provides evidence of truly mental operations, thereby meeting Piaget's original criteria for the presence of 'true' analogical reasoning.

From the picture analogy game, we know that the ability to reason by analogy is present by at least age 4. However, the analogy game may **still** have underestimated analogical ability. This is because relational familiarity was not measured **independently** of analogical success. Instead, it was simply assumed that familiar relations had been selected for the analogies, leaving open the possibility that the younger children may have failed in some trials because the relations used in those particular analogies were unfamiliar to them. Alternatively, some children may have failed some analogies because they were actually reasoning about relations that were **different** from those intended by the experimenter — like Lucas.

THE RELATIONSHIP BETWEEN RELATIONAL KNOWLEDGE AND ANALOGICAL RESPONDING

The idea that children's analogical performance depends on their relational knowledge has been called the *'relational familiarity' hypothesis*. In order to establish whether children's use of analogical reasoning is knowledge-based, dependent on relational familiarity rather than analogical ability, relational knowledge **as well as** analogical ability needs to be assessed. This can be done by changing the control task in the picture matching game. The appropriate control task measures children's knowledge of the relations being used in the analogies that are presented in the item analogy task.

A second set of analogy experiments using the picture matching game were thus carried out to test the relational familiarity hypothesis. This time, item analogies based on physical causal relations like *melting*, *cutting* and *wetting* were used. These relations are acquired early in development, between 3 and 4 years of age. Children were given analogies like '*chocolate is to melted chocolate as snowman is to* ?', and '*playdoh is to cut playdoh as apple is to* ?'. Knowledge of the causal relations required to solve the analogies was measured by giving the children pictures of items that had been causally transformed (e.g., cut playdoh, cut bread, cut apple), and asking them to select the causal agent responsible for the change from a set of pictures of possible agents (e.g., a knife, water, the sun).

This 'causal relations' version of the picture matching game was given to children aged 3, 4 and 6 years of age. The results showed that both analogical success and causal relational knowledge increased with age. The 3-year-olds solved 52% of the analogies and 52% of the control sequences, the 4-year-olds solved 89% of the analogies and 80% of the control sequences, and the 6-year-olds solved 99% of the analogies and

100% of the control sequences. There was also a significant **conditional** relationship between performance in the two conditions, as would be predicted by the relational familiarity hypothesis. This conditional relationship showed that individual children's performance in the analogy task was intimately linked to those individual children's knowledge of the corresponding causal relations. Analogical success had thus been shown to be highly dependent on relational knowledge. These experiments showed that Piaget's theory of analogical development could no longer be upheld. If analogy is one of the basic cognitive processes underlying intellectual development, then it should be found at work in many other areas of cognition.

ANALOGIES IN COGNITIVE DEVELOPMENT

Analogies in Piagetian Tasks

An elegant theory of how analogical reasoning may contribute to performance in Piagetian logical tasks has been proposed by Halford (1993). Halford's basic claim is that much logical reasoning is analogical. According to his theory, children can use representations of everyday relational structures as a basis for analogies to new, isomorphic problems that share the same relational structures. For example, in order to solve a Piagetian transitive inference problem of the form *Tom is happier than Bill, Bill is happier than John, who is happiest?* a child can use an analogy from a familiar ordered stucture that may already be represented in memory. An example is the ordering structure **A above B above C.** Halford has suggested that all of Piaget's logical tasks that are characteristic of the 'concrete operational' stage of logical development (transitive reasoning, class inclusion, conservation) require analogical mappings based on pairs of relations.

In order to test the idea that Piagetian 'concrete operational' tasks can be solved by using appropriate analogies, therefore, we must first examine children's ability to map pairs of relations. This can be done by extending the classical analogy task by linking the A and B terms by two relations rather than one. Goswami, Leevers, Pressley and Wheelwright (1998) designed a set of analogies based on pairs of physical causal relations, extending the technique used by Goswami and Brown (1989). We asked 3-, 4-, 5- and 6-year-old children to make relational mappings based on either single causal relations like **cut**, **paint,** and **wet**, or pairs of causal relations, like **cut + wet** and **mend + paint**. This experimental paradigm provides a relatively pure test of the ability to make analogies about pairs of relations.

Our experiment had four conditions, a single-relation analogy condition (e.g., *apple: cut apple:: hair: cut hair*), a double-relation analogy condition (e.g., *apple: cut, wet apple:: hair: cut, wet hair*), a single-relation control condition and a double-relation control condition. In the control conditions, the children were asked to select the picture of the causal agent or the pair of causal agents responsible for the causal changes shown in the analogies, following Goswami and Brown (1989).

Children's performance in the analogy and the control conditions was then examined as a function of Condition and Age. The pattern of the results was remarkably similar to the pattern found in the causal relations analogies used by Goswami and Brown (1989). There was a close correspondence between analogy performance and performance in the relational knowledge control conditions for both the single relation and the double relation analogies. For the single relation conditions, the 3-year-olds solved 33% of the analogies and 46% of the control sequences, the 4-year-olds solved 51% of the analogies and 63% of the control sequences, the 5-year-olds solved 72% of the analogies and 76% of the control sequences, and the 6-year-olds solved 89% of the analogies and 88% of the control sequences. For the double relation conditions, the 3-year-olds solved 13% of the analogies and 31% of the control sequences, the 4-year-olds solved 50% of the analogies and 50% of the control sequences, the 5-year-olds solved 62% of the analogies and 74% of the control sequences, and the 6-year-olds solved 78% of the analogies and 91% of the control sequences. Analyses demonstrated no interaction between age and number of relations, although the main effect of number of relations almost reached significance, reflecting the fact that children of all ages found the double relation analogies and control sequences more difficult than the single relation analogies and control sequences. Goswami et al. concluded that the ability to solve analogies based on pairs of relations was governed by relational familiarity. As long as familiar relational structures are chosen as a basis for analogy, therefore, young children should be able to use analogies to help them to solve Piagetian reasoning tasks.

Analogies in a Transitive Mapping Task

Halford has suggested that familiar ordered structures may provide useful analogies for transitive reasoning tasks. Family members provide a familiar example of an ordering structure based on size, as in most families the father (F) is taller than the mother (M), and the mother is taller than the young child (C). If knowledge of the familiar relational structure $F > M > C$ is present in young children, then children who have mentally represented this relational structure should be able to solve transitive mapping tasks using less familiar relations.

Goswami (1995) examined this hypothesis using Goldilocks and the Three Bears as a familiar example of family size relations (Daddy Bear > Mummy Bear > Baby Bear). Three- and 4-year-old children were asked to use the relational structure represented by the Three Bears as a basis for solving transitive ordering problems involving perceptual dimensions such as temperature, loudness, intensity, and width. The transitive mapping test was presented by asking the children to imagine going to the Three Bears' house, and then to imagine looking at their different belongings. This imagination task constituted a fairly abstract test. For example, the imaginary bowls of the Three Bears' porridge could be either **boiling hot, hot,** or **warm**, and the child had to decide which bowl of porridge belonged to which bear. In order to give the correct answer, the child had to map the transitive height ordering of Daddy, Mummy, and Baby Bear to the different porridge temperatures, giving Daddy Bear the boiling hot porridge, Mummy Bear the hot porridge, and Baby Bear the warm porridge (these mappings do not follow the original fairy tale, in which Daddy Bear's porridge was too salty, and Mummy Bear's was too sweet).

The results showed that the percentage of correctly ordered mappings approached ceiling for the 4-year-olds for most of the dimensions used. The lowest levels of performance occurred for **width** (of beds, 62% correct), and **hardness** (of chairs, 76% correct), and the highest occurred for **temperature** (of porridge, 95% correct). Performance with the width dimension (wide bed, medium bed, narrow bed) was possibly affected by worries that a baby could fall out of a narrow bed, as many children allocated the medium bed to Baby Bear. They were then left without a bed for Mummy Bear. The 3-year-olds produced correctly ordered mappings for only some of the dimensions, performance being above chance (17%) for the dimensions of temperature of porridge (31% correct), pitch of voice (31% correct), and height of mirrors (62% correct, but an isomorphic relation). Relational familiarity and real-world knowledge about family size relations seem to have helped the 3-year-olds with these particular dimensions. The children are unlikely to have based their correct mappings on the story, as none of these dimensions was mentioned in the **Three Bears** book that was read to them as part of the study.

Analogies in a Class Inclusion Task

Families also provide a familiar example of an inclusive relationship, as family members can be divided into two distinct sub-sets, parents and children, both of which are members of the total set of family members (Halford, 1993). In order to see whether the family as a familiar example of inclusive relations could act as a basis for successful performance in Piagetian class inclusion tasks, Goswami, Pauen and Wilkening (1996) devised the 'create-a-family' paradigm. In this paradigm, children were shown a toy family, for example a family of toy mice (2 large mice as parents, 3 small mice as children). Their job was to create analogous families (2 patents and 3 children) from an assorted pile of toys (such as toy cars, spinning tops, balls and helicopters). After the children had correctly created 4 analogous families, they were given 4 class inclusion problems involving toy frogs, sheep, building blocks and balloons. The class inclusion problems were posed using collection terms ('group', 'herd', 'pile', 'bunch'). The children in Goswami et al.'s study (4- to 5-year-olds) had all failed the traditional Piagetian class inclusion task, which was given as a pretest ("Are there more red flowers or more flowers?"). A control group of children received the same class inclusion problems using collection terms, but did not receive the 'create-a-family' analogy training session.

Goswami et al. found that more children in the 'create-a-family' analogy condition than in the control condition solved at least 3 of the 4 class inclusion problems involving frogs, sheep, building blocks and balloons. This effect was particularly striking at age 4, in which no improvement at all was found in the control group with the collection term wording. It should be remembered that all of the children had previously failed Piagetian class inclusion tasks. Goswami et al. argued that this improvement was a result of the use of analogies based on a representation of family structure.

Analogies in Foundational Domains

One popular view of cognitive development is that conceptual development can be understood in terms of three 'foundational' domains. These are the domains of naive biology, naive physics, and naive psychology (Wellman & Gelman, in press). Wellman and Gelman argue that, rather than developing a monolithic understanding of the world, young children develop distinct conceptual frameworks to describe these 'foundational' domains, even though many concepts will be represented in more than one of these foundational frameworks (for example, persons are psychological entities, biological entities **and** physical entities). Wellman and Gelman suggest that children will use at least two levels of analysis within any framework, one that captures surface phenomena (mappings based on attributes) and another that penetrates to deeper levels (mappings based on relations). This means that analogies should be at work within foundational domains. Although no-one has yet studied the role of analogies in the foundational domain of psychology ('theory of mind'), studies of the role of analogies in developing conceptual understanding in the domains of naive biology and naive physics can be found.

Analogy as a Mechanism for Understanding Biological Principles

Evidence that analogy is an important mechanism for understanding biological principles comes from a series of studies by Inagaki and her colleagues. They were interested in how often children would base their predictions about biological phenomena on analogies to people: the 'personification' analogy. As human beings are the biological kinds best known to young children, it seems plausible that children may use their biological knowledge about people to understand biological phenomena in other natural kinds. For example, Inagaki and Sugiyama (1988) asked 4-, 5-, 8- and 10-year-olds a range of questions about various properties of 8 target objects, including "Does x breathe?", "Does x have a heart", "Does x feel pain if we prick it with a needle", and "Can x think?". The target objects were people, rabbits, pigeons, fish, grasshoppers, trees, tulips and stones. Prior similarity judgements had established that the target objects differed in their similarity to people in this order, with rabbits being rated as most similar and stones being rated as least similar. The children all showed a decreasing tendency to attribute the physiological properties ("Does x breathe") to the target objects as the perceived similarity to a person decreased. Apart from the 4-year-olds, very few children attributed physiological attributes to stones, tulips and trees, and even 4-year-olds only attributed physiological properties to stones 15% of the time. A similar pattern was found for the mental properties ("Can x think?"). This study supports the idea that preschoolers' understanding of biological phenomena arises from analogies based on their understanding of people.

Analogy as a Mechanism for Understanding Physical Principles

Evidence that analogy is an important mechanism for understanding physical principles comes from a series of studies by Pauen and her colleagues. Pauen has studied children's understanding of the principles governing the interaction of forces, by using a special apparatus called the 'force table'. The force table consists of an object that is fixed at the centre of a round platform. Two forces act on this object, both represented by plates of weights. The plates of weights hang from cords attached to the central object at either 45', 75' or 105' to each other. The children's job is to work out the trajectory of the object once it is released from its fixed position. Their predictions concerning this trajectory are scored in terms of whether they consider only a single force (plate of weights), or whether they integrate both forces in order to determine the appropriate trajectory. The force table problem is presented to the children in the context of a story about a King (central object) who has got tired of skating on a frozen lake (the platform) and who wants to be pulled into his royal bed on the shore. Children aged 6, 7, 8 and 9 years of age were tested.

Pauen found that most of the younger children (80 - 85%) predicted that the king would move in the direction of the stronger force only (the larger plate of weights). An ability to consider the two forces simultaneously was only shown by some of the 9-year-olds (45%). Such integration rule responses were shown by the majority of the adults tested (63%). Pauen speculated that this may have been because the children who received the plates of weights applied a balance scale analogy to the force integration problem. A balance scale analogy gives rise to one-force-only solutions, which are incorrect.

This idea about the balance scale analogy was prompted by the comments of the children themselves, who said that the force table reminded them of a balance scale (presumably because of the plates of weights). This led Pauen to propose that the children were using spontaneous analogies in their reasoning about the physical laws underlying the force table, analogies that were in fact misleading. To investigate this idea further, Pauen and Wilkening (in press) gave 9-year-old children a training session with a balance scale prior to giving them the force table problem. One group of children received training with a traditional balance scale, in which they learned to apply the one-force-only rule, and a second group of children received training with a modified balance scale that had its centre of gravity below the axis of rotation (a 'swing boat' suspension). This modified balance scale provided training in the integration rule, as the swing boat suspension meant that even though the beam rotated towards the stronger force, the degree of deflection depended on the size of **both** forces.

Following the balance scale training, the children were given the force table task with the plates of weights. A third group of children received only the force table task, and acted as untrained controls. Pauen and Wilkening argued that an effect of the analogical training would be shown if the children who were trained with the traditional balance scale showed a greater tendency to use the one-force-only rule than the control group children, while the children who were trained with the modified balance scale showed a greater tendency to use the integration rule than the control group children. This was exactly the pattern that they found. The children's responses to the force table problem varied systematically with the solution provided by the analogical model. These results suggest that the children were using spontaneous analogies in their reasoning about physics, just as we have seen them do in their reasoning about biology.

REFERENCES

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.) *Similarity and analogical reasoning*, (pp. 199-241). London: Cambridge University Press.

Goswami, U. (1995). Transitive relational mappings in 3- and 4-year-olds. *Child Development, 66*, 877.

Goswami, U. & Brown, A. (1989) Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition, 35*, 69-95.

Goswami, U. & Brown, A.L. (1990). Higher-order structure and relational reasoning. *Cognition, 36*, 207-226.

Goswami, U. et al., (1998). Causal reasoning about pairs of relations and analogical reasoning in young children. *British Journal of Developmental Psychology.*

Goswami, U., Pauen, S., & Wilkening. F. (1996). *The effects of a 'family' analogy in class inclusion tasks*. Manuscript in preparation.

Halford, G.S. (1993). *Children's Understanding: The Development of Mental Models*. Hillsdale, NJ: Erlbaum.

Inagaki, K., & Sugiyama, K. (1988). Attributing human characteristics. *Cognitive Development, 3*, 55-70.

Pauen, S., & Wilkening, F. (in press). Children's spontaneous use of analogies in explaining natural phenomena. *J. of Experimental Child Psychology.*

Polya, G. (1957). *How to Solve It*. Princeton: University Press.

Wellman, H.M., & Gelman, S.A. (in press). Knowledge acquisition in foundational domains. D. Kuhn & R. Siegler (Eds), *Handbook of Child Psych, 2*.

RELATIONAL PROCESSING IN HIGHER COGNITION: IMPLICATIONS FOR ANALOGY, CAPACITY AND COGNITIVE DEVELOPMENT

Graeme S. Halford
University of Queensland School of Psychology University of Queensland Brisbane 4072, Australia Tel: 61 7 3365 6401 (w) 61 7 3844 0183 (h) Fax: 61 7 3365 4466
e-mail: gsh@psy.uq.oz.au

William H. Wilson
University of New South Wales

Steven Phillips
Electrotechnical Laboratory

ABSTRACT

It is proposed that models based on processing relations capture the structure sensitivity of higher cognitive processes while they can also be compared with more basic processes such as associations. Relations have the following properties that are not shared by associations: there is an explicit symbol for each relational instance, allowing it to be manipulated, higher-order relations can be formed that have lower-order relations as arguments, given any N-1 components of an n-ary relation the remaining component can be retrieved (omni-directional access), and representation of relational instances is a prerequisite to analogical mapping. A model is proposed in which each component of a relational instance is represented by a vector, and the binding is represented by computing the outer product of the vectors. This architecture has been used to model analogy and human memory. It can also be used to model structural effects on both similarity and category formation. Computational cost increases exponentially with representational rank, defined as number of components that are bound into a representation. Thus the model provides a natural explanation for processing capacity limitations in humans and higher animals. Each rank corresponds to a class of psychological processes, neural nets, and empirical criteria. The ranks and typical concepts which belong to them, are: Rank 0, elemental association; Rank 1, content-specific representations and configural associations; Rank 2, unary relations, class membership, variable-constant bindings; Rank 3, binary relations, proportional analogies; Rank 4, ternary relations, transitivity and hierarchical classification; Rank 5, quaternary relations, proportion and the balance scale. Rank 6, quinary relations. Rank 0 can be performed by 2-layered nets, rank 1 by 3-layered nets, and ranks 2-6 by tensor products of the corresponding number of vectors. All animals with nervous systems perform rank 0, vertebrates perform rank 1, other primates perform rank 2-3, but ranks 4-6 are uniquely human. Rank also increases with age. Implications of this model are developed for human reasoning and cognitive development.

In this paper we will present an outline of a theory that provides a general metric for cognitive complexity, and specifies properties of higher cognitive processes in a way that enables them to be distinguished systematically from more basic cognitive functions. The theory distinguishes the cognition of humans from other animals, distinguishes levels of cognitive development, and accounts for processing loads in cognitive tasks, within a common metric based on structural complexity. The levels of complexity are related systematically both to neural net architectures and to empirical criteria. Analogy has a central role in this theory, first because it is a core mechanism in higher cognition, and second because lower cognitive processes cannot implement analogy.

Although interest in analogy dates back to near the beginning of scientific psychology (Piaget, 1950; Spearman, 1923) understanding of human analogical reasoning accelerated dramatically in the 1980s (Gentner, 1983; Gick & Holyoak, 1983). Analogy is a natural mechanism for human reasoning, but we will suggest that its involvement in higher cognition might be even greater than previously realised. It has proven difficult to produce effective models of human reasoning based on logical inference rules. Such models do exist (Braine, 1978; Rips, 1989) but most theorists have chosen to model reasoning on the basis of alternative psychological mechanisms such as memory retrieval (Kahneman & Tversky, 1973) mental models (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) or pragmatic reasoning schemas (Cheng & Holyoak, 1985). Analogy can play a role in a human reasoning and is also entailed in some significant ways with a number of other models. We can illustrate this using pragmatic reasoning schemas.

Although it has become fashionable to interpret pragmatic reasoning schemas as being specialised for deontic reasoning, they may be more widely applicable. Consistent with this, we will use the definition of pragmatic reasoning schemas as structures of general validity that are induced from ordinary life experience. One type of pragmatic reasoning schema, permission, is known to improve performance on the Wason Selection Task (Cheng & Holyoak, 1985). In this task participants are given four cards containing $p, \bar{p}, q, \bar{q}$ and asked which cards must be turned over to test the rule p -> q. Analogy plays a central role here, because as Figure 1 shows, the elements and relations presented in the WST task can be mapped into a permission or prediction schema. This can be done by application of the principles that are incorporated in contemporary computational models of analogy (Falkenhainer, Forbus, & Gentner,

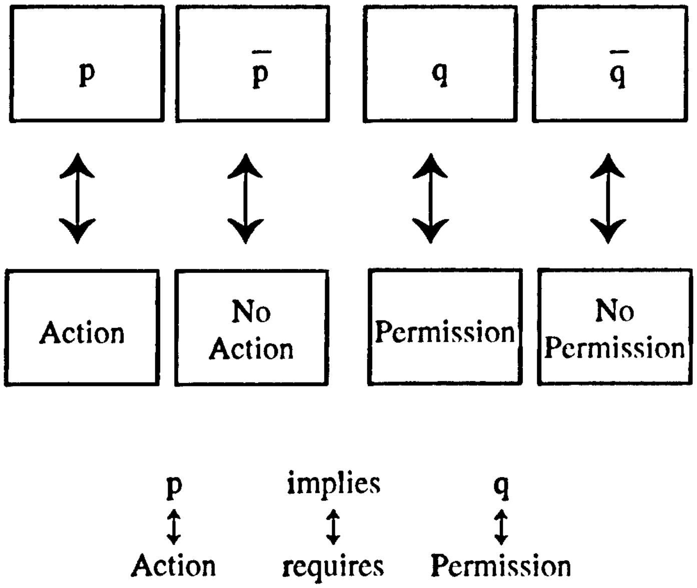


Figure 1.

1989; Gray, Halford, Wilson, & Phillips, 1997; Hummel & Holyoak, 1997; Mitchell & Hofstadter, 1990) and no special mechanism is required.

Possible reason why induction of a permission schema improves performance is that, as Table 1 shows, permission is isomorphic to the conditional. Extending this argument, a possible reason for the tendency to respond in terms of the biconditional p <-> q, is that participants may otherwise interpret the rule as a prediction. As Table 1 shows, prediction is isomorphic to the biconditional. This argument has been presented in more detail elsewhere (Halford, 1993). It implies that the importance of permission is not that it is deontic, but that it is isomorphic to implication. While we would not suggest that this argument does justice to the extensive literature on either the Wason Selection Task or pragmatic reasoning schemas, it does serve to illustrate that analogy can serve as the basic mechanism even in tasks such as WST that might normally be considered to entail logical reasoning.

ANALOGY, RELATIONS AND HIGHER COGNITIVE PROCESSES

Although there are big differences between contemporary computational models of analogy, there is some degree of consensus about the core processes. In particular, it seems clear that analogy is a matter of mapping relations or relational instances between two representations. The core principles seem to be: the elements in one structure, the *base* are mapped uniquely to the elements of the other structure, the *target*, and; if a predicate P in the base is mapped to the predicate P' of the target, the arguments of P are mapped to the arguments of P'. The relational instances may be coded in the input (Gray et al., 1997; Hummel & Holyoak, 1997, Falkenhainer, 1989 #1136) or they may be constructed dynamically during the running of the model (Mitchell & Hofstadter, 1990) but a mapping between relational instances seems to constitute the essence of analogy in most models. It seems fair to say that an organism that could not represent relations or relational instances could not perform analogy. If we accept that analogy is one of the core processes in higher cognition, then ability to process relations and relational instances is also likely to be important in higher cognition. This is really an argument for the importance of structure in higher cognition, because relations are the essence of structure (a structure is a set on which one or more relations is defined).

Our next step is to consider those properties of higher cognitive processes on which there seems to be reasonable consensus. One such property is representation of structure, together with ability to operate on that structural representation. This is generally seen as the essence of higher cognition. The central role of structure in higher cognitive processes has been recognised historically (Humphrey,

			Permission Schema (Symbolic)			*Conditional*			*Biconditional (Prediction)*		
Permission schema		*Action → Permission*			*A → P*	*A*	*P*	*A → P*	*p*	*q*	*p ↔ q*
Action	permission	allowed	A	P	+	1	1	1	1	1	1
Action	no permission	not allowed	A	$\bar{P}$	-	1	0	0	1	0	0
No action	permission	allowed	$\bar{A}$	P	+	0	1	1	0	1	0
No action	no permission	allowed	$\bar{A}$	$\bar{P}$	+	0	0	1	0	0	1

Table 1.
The Structure of the Permission Schema

1951) and by a number of writers in this century, including Gestaltists (Wertheimer, 1945), Piagetians (Piaget, 1950), information processing theorists (Anderson, 1983; Hunt, 1962; Miller, Galanter, & Pribram, 1960; Newell, 1990) and linguists (Chomsky, 1980; Fodor, 1975). One role of analogy is to form mappings between structures, so on these grounds also analogy might be considered a core process in higher cognition.

There is also reasonable consensus that higher cognitive processes entail variables, which are essential to the generality and content-independence that characterise higher cognitive processes.. An entire generation of cognitive models are based on rules, a distinguishing characteristic of which is that they relate variables. Production rules are perhaps the most common example (Anderson, 1983; Newell, 1990) and production systems normally have provision for variable binding. Smith, Langston and Nisbett (1992) make a case for logical inference rules being used in natural reasoning. These rules relate variables. For example, modus ponens is a logical inference rule of the form *if p then q, p therefore q*, where *p* and *q* are variables. Pragmatic reasoning schemas (Cheng & Holyoak, 1985) are more content-specific than abstract logical inference rules, but still relate variables. Thus the permission schema can be expressed as: to perform act *a*, you must have permission *p*.

Analogies can simulate variables by putting instances of a relation in correspondence with each other. Consider for example the following relational instances:

larger(whale,fish),
larger(horse,dog),
larger(5,3).

Each relational instance has two roles or slots, one filled by the larger entity and one by the smaller entity in a given pair. Because each role can be instantiated in a variety of ways, it effectively functions as a variable, but only if the arguments are in correspondence. It would not be true if the relational instances were cross-mapped, as in this case:

larger(whale,fish)
larger(dog,horse)

Models of analogy include mechanisms for ensuring structural correspondence. Indeed this is a core process in analogy models. Therefore they provide a mechanism that is capable of at least limited processing of variables.

Higher cognitive processes are widely regarded as incorporating symbols, even though the issue has become complicated by the debate between proponents of symbolic and connectionist models. Newell argued that symbols and a system that operates on them are necessary for intelligent action (see Newell, 1990, p. 170). Fodor and Pylyshyn (1988) argue that symbols are vital to cognition. Smolensky (1988), a connectionist modeller, does not deny the importance of symbols *per se*, but seeks to explain them at the subsymbolic level, rather than accepting them as a primary datum. With this proviso, there does seem to be widespread acceptance of the importance of symbols in higher cognitive processes.

Analogical reasoning mechanisms operate on relations that are symbolic in the sense that they include a label that specifies the link between the entities that are related. Thus in the instances considered above, the entities in the pairs (whale,fish) and (horse,dog) are linked by the relation symbol "larger". Mathematically, an n-ary relation is a subset of the cartesian product of *n* sets, but the subset is typically specified by a label; for example, >{. . (3,2), . . , (5,1), . . ,}. The existence of a label and an ordering over relational elements (i.e., R(a,b) is not the same as R(b,a)) are important characteristics that distinguishes relations from other psychological structures such as associations, as we will argue later. We will briefly consider some further properties of higher cognition.

Compositionality has come to be accepted as a property of higher cognitive processes since the work of Fodor and Pylyshyn (1988). In essence it means that the components of a cognitive representation retain their identity when they are composed into more complex representations, and both the components and the composites are semantically evaluable. As we will see, there are cognitive processes such as configural association, for which these properties do not hold.

Systematicity is another property that has been accepted as important in higher cognition since Fodor & Pylyshyn (1988) although it too has been the subject of some controversy (Niklasson & van Gelder, 1994; Van Gelder & Niklasson, 1994).In essence, it implies generalisation to all logically or structurally equivalent situations, although it is generally accepted that content can also influence performance, independent of structure, to some extent. Analogy clearly has the potential to be a core mechanism in achieving systematicity.

Categories are another property of higher cognition. We will not consider this complex topic here, except to say that categories must entail a label that is independent of content.

Modifiability on line is a property of higher cognitive processes that has been highlighted by the work of Clark and Karmiloff-Smith (1993). Higher cognitive processes should also be productive or generative, in the sense that they can produce or comprehend new sentences, can generate new representations, and make new inferences. This is true of both human and nonhuman primates, because apes show some inventiveness (Kohler, 1957) and ability to draw inferences (Tomasello & Call, 1997). Furthermore, though we will not pursue the question here, the approach we have adopted can model some limited forms of creativity (Halford, Wiles, Humphreys, & Wilson, 1993).

We do not include conscious awareness and language as criterial properties of higher cognition. Awareness has proven to be a difficult criterion to use, as the implicit learning literature has shown (Neal & Hesketh, 1997). As we wish to include some nonlinguistic, nonhuman species as having at least some forms of higher cognition, then language cannot be included either. We see conscious awareness and language as correlated rather than criterial properties of higher cognition.

We want to suggest that relational processing can capture the properties of higher cognition. Relations are preferable to rules, which have been used to model higher cognitive processes and to distinguish them from basic processes that have been characterised as associative (Sloman, 1996) or instance-based (Smith et al., 1992). Some cognitive representations such as loves(John,Mary) or contains(cup,drink) are not rules, but can be expressed as relations. The concept of n-ary relation is general enough to express any rule, it has the advantage of a precise mathematical definition, and effects of relational complexity on processing load are known -a{Blank or a = BBS, Which one b??}(Halford et al., in press). Relations are increasingly being utilised as the basis for models of higher cognitive processes. In addition to analogy, the importance of relational processing has been recognised in similarity (Markman & Gentner, 1993), induction (Lassaline, 1996), and categorisation (Medin, 1989). Mental model theory, which can now account for a wide range of phenomena in human reasoning (Gentner & Stevens, 1983; Halford, 1993; Johnson-Laird, 1983; Polk & Newell, 1995), is based on representation of relations between entities. Phillips, Halford, and Wilson (1995) have argued that the associative-relational distinction can subsume the implicit-explicit distinction of Clark and Karmiloff-Smith (1993). Propositions, which are the core of some models of higher cognitive processes, can be treated as relational instances (Halford, Wilson, & Phillips, in press), section 2.2.2). For example the proposition loves(Joe,Jenny) is a relational instance.

Another big advantage of relations is that they can be compared directly with associations. The importance of this is that association has been accepted as a fundamental process in psychology virtually throughout the history of the discipline, and even many contemporary models incorporate it in one form or another. Therefore it is a disadvantage for associative and cognitive models to exist in conceptual worlds that do not communicate. We will suggest that basic processes, such as association, and higher cognitive process, which we identify with relations, can be incorporated into an overarching theory that integrates psychological processes at all levels. First however we will consider the properties of associative and relational knowledge in more detail.

ASSOCIATION

By contrast with higher cognition, association is not seen as inherently structural (Fodor & Pylyshyn, 1988; Humphrey, 1951). It differs from relational knowledge in a number of critical ways, one of which is that it is not symbolic. To illustrate, let us consider two commonplace relations, between cup and drink and between cup and saucer: i.e. contains(cup,drink) and placed-on(cup,saucer). The relation-symbols (or predicates) *contains* and *placed-on* specify the type of link represented, containment or superposition. Contrast this with associations; cup is associated with drink, and cup is associated with saucer. The associations *per se* do not specify the relations between cup and drink, or between cup and saucer, nor do these associations *per se* capture the fact that the relations are quite different. It is easy to overlook this because we know that a cup *contains* a drink and that a cup is *placed on* a saucer, so we tend to see this information in the associative link. The associative link is causal but does not capture the structure (Fodor & Pylyshyn, 1988). Associative links are unlabelled and all of the same kind, differing only in strength (Humphrey, 1951). The need for labelled links has been recognised however in models of higher cognitive structures such as propositional networks, in which links between nodes carry labels such as "agent", "object", "location". An explicit symbol for a link therefore appears to be a property that distinguishes relational from associative processes. Our aim now is to define the properties of relational processes so that they capture the essence of higher cognition and can be compared directly with association.

PROPERTIES OF RELATIONAL PROCESSES

A relation that relates *n* entities, or **n-ary relation** is a subset of the cartesian product of *n* sets: i.e. $R(a_1,a_2,\ldots,a_n)$ is a subset of $S_1 x S_2 x \ldots x S_n$. A relation is identified by the relation symbol, R, and the entities by argument symbols, $a_1,a_2,\ldots,a_n$. For example the relation "larger" identifies a specific subset of a cartesian product, that subset in which the first entity is always larger than the second; i.e. $a_1 > a_2$. There must be a binding between entities and arguments which preserves the truth of the relation; thus contains(cup,drink) is true but contains(drink,cup) is not.

Symbolisation, or an explicit label specifying the link, is a property of relations, but not of associations.

Higher-order relations have lower-order relations as arguments; e.g. in cause(shout-at(Tom,John), hit(John,Tom)) cause is a higher-order relation, with shout-at(Tom,John) and hit(John,Tom) as arguments.

Systematicity means that relations imply other relations, and can be captured by higher-order relations; e.g. >(a,b) implies <(b,a), can be written as the higher-order relation implies(>(a,b),<(b,a)).

Association does not share these properties. Associations can be chained, so that the output of one association is the input to another: $E_1 \rightarrow E_2 \rightarrow E_3 \ldots\ldots \rightarrow E_n$, and may converge, so that E_1 and E_2 elicit E_3, or diverge, so that E_1 elicits E_2 and E_3 However associations are not identified by a symbol, and the associative link *per se* cannot be an argument to another association. Therefore the recursive, hierarchical structures that can be formed using higher-order relations do not appear to be possible with associations.

Compositionality means that the components of the relation, symbol and arguments, retain their identity when bound into a structure; e.g. in larger(whale,dolphin), the components "larger", "whale" and "dolphin" retain their identity when bound into the relation. This is not inherent in association, as we will see when we consider configural associations.

Modifiability by strategic processes, without information input, is possible for relations, whereas associations are modified incrementally on the basis of experience.

Omni-directional access means that, given all but one of the components of a relational instance, we can access (i.e. retrieve) the remaining component. For example, given the relational instance mother-of(woman,child), and given mother-of(woman,?) we can access "child", whereas given mother-of(?,child) we can access "woman", and given ?(woman,child) we can access "mother-of". Although backward association may be possible, omni-directional access does not appear to be inherent in association.

Complexity can be defined by the "arity" or number of arguments of a relation (Halford et al., 1994; Halford et al., in press). Each argument corresponds to a source of variation or dimension, so an *n*-nary relation is a set of points in *n*-dimensional space. Dimensionality is related to processing load. **Capacity** is limited by the number of dimensions (or number of interacting variables) that can be processed in parallel. Data in the literature, and from our own laboratory, indicates quaternary relations (Rank 5) are the most complex that can be processed in parallel by most humans. Concepts too complex to be processed in parallel are handled by *segmentation* (decomposition into smaller segments that can be processed serially) and *conceptual chunking* (recoding representations into lower rank, but at the cost of making some relations inaccessible). For example, velocity = distance/time, is a ternary relation, and is Rank 4, but can be recoded to rank 2, a binding between a variable and a constant (Halford et al., in press), Section 3.4.1). Difficulty can vary because of factors other than capacity, including declarative and procedural knowledge and amount of iteration (e.g. constructing a 5-term series from premises a>b, b>c, c>d, d>e requires the integration process to be iterated 3 times; a>b, b>c yields a,b,c, then this is integrated with c>d to yield a,b,c,d, etc.).

In the next section we argue that each level of cognitive functioning can be assigned to an equivalence class of equal structural complexity, and that the classes can be ordered according to their complexity. They are ordered according to representational rank, defined as the number of components in cognitive representations, given that the components retain their identity when bound into more complex representations. An important feature of this idea is that the ranks correspond across the three domains of psychological process, neural net structure, and empirical observation. Each rank corresponds to a class of neural net architectures and can be identified by specific empirical criteria. It is an extension of a theory that defines processing capacity in terms of relational complexity (Halford et al., in press).

REPRESENTATIONAL RANK

Representational rank corresponds to the number of components of a representation, given that the components retain their identity when bound in a more complex representation. The metric is shown in Figure 2, together with corresponding psychological processes and neural net architectures. The metric combines relational complexity with two nonstructural levels, elemental and configural association, enabling the basic properties of all levels of cognition to be defined within a single system. Rank = *n*+1 where *n* is the dimensionality or arity of a relation. We will now give an overview of the ranks.

Figure 2.**Rank** 0 corresponds to **Elemental associations**, which comprise links between pairs of entities: $E_1 \rightarrow E_2$

They are Rank 0 because there is no representation other than input and output, and they can be implemented by 2-layered nets. In principle Rank 0 can be assessed by any associative learning test, and because ability to perform at this level is not in question for vertebrates, or even for most invertebrates, no special assessment is intended.

Rank 1 corresponds to **Configural associations**, in which one cue is modified by another. They have the form: $E_1,E_2 \rightarrow E_3$ An example is conditional discrimination, shown in Table 2. This cannot be acquired through elemental association, because of associative interference (each element, colour or shape, is equally associated with each outcome). They can be learned by fusing or "chunking" elements into a configuration such as "black/triangle". This avoids associative interference but at the cost that the components lose their identity, (e.g. "triangle" is not the same in "black/triangle" as in "white/triangle") so the structure of the task is not represented. Thus the representation is holistic and nonstructural. Configural learning cannot be implemented by 2-layered nets (Minsky & Papert, 1969; note that conditional discrimination is isomorphic to exclusive-OR). They can be implemented with three-layered nets, by using units in the hidden layer to represent configurations of features such as "black&triangle" (Schmajuk & DiCarlo, 1992).

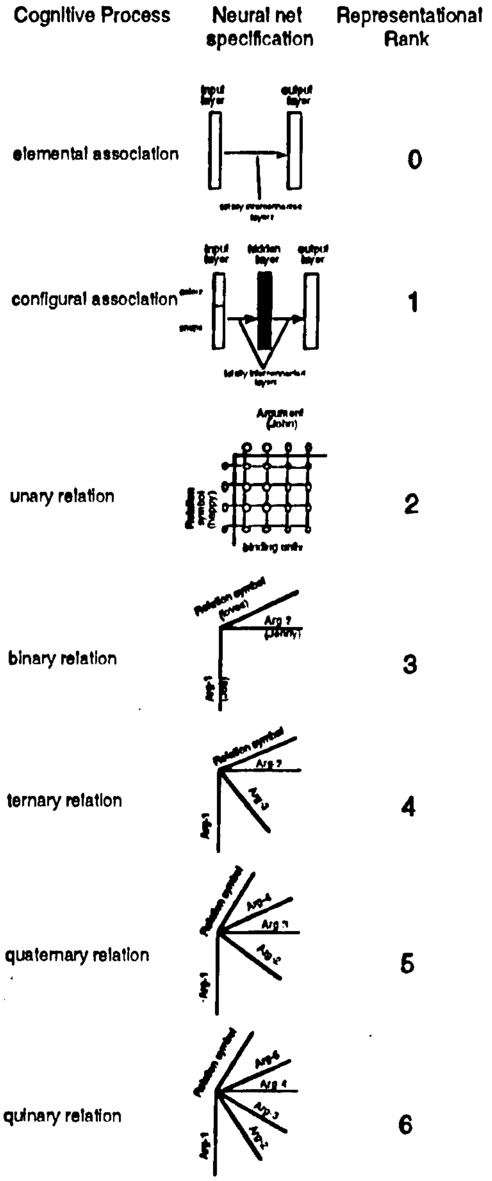

Cognitive Process
Neural net specification
Representational Rank
elemental association
input layer
output layer
0
configural association
input layer
hidden layer
output layer
1
unary relation
Argument (John)
binding units
2
binary relation
Arg 2
3
ternary relation
Arg 2
Arg 1
4
quaternary relation
Arg 3
Arg 1
5
quinary relation
Arg 4
Arg 3
Arg 1
6

Figure 2. Ranks 0-6, with schematic neural nets. Input and output layers are omitted for Ranks 2-6.

Ranks 2-6 are structural, and complexity increases with rank. We will consider the main properties of each rank.

Rank 2 corresponds to **unary relations** which are a binding between a relation symbol and an argument symbol. An example would be the proposition happy(John). Indicators of Rank 2 include symbolic representation of categories and understanding word reference.

Rank 3 corresponds to binary **relations**, which represent common states and actions in the world, such as larger(whale,dolphin), or loves(Joe,Jenny).

Rank 4 corresponds to **Ternary relations** such as "love-triangle", which is a relation between three people. They can be interpreted as bivariate functions, and binary operations. For example, the binary operation of arithmetic addition consists of the set of ordered triples of +{. . , (3,2,5), . . , (5,3,8), . . , . . } and is a ternary relation. Many cognitive tasks that cause difficulty for young children, including transitivity and class inclusion, are ternary relations (Halford, 1993; Halford et al., in press).

Rank 5 corresponds to **quaternary relations**. Proportion, a/b = c/d, is a quaternary relation. Comparison of moments on the balance scale (Siegler, 1981) is another example.

Table 2. Conditional discrimination, with isomorphic transfer task.

Original task			Transfer task		
black	triangle →	R+	green	circle →	R+
black	square →	R-	blue	cross →	?
white	triangle →	R-	green	cross →	?
white	square →	R+	blue	circle →	?

Rank 6 corresponds to **quinary relations**. Some complex reasoning tasks, such as categorical syllogisms and meta logical tasks, require Rank 6.

NEURAL NET MODELING OF REPRESENTATIONAL RANKS

Neural nets can be rank-ordered according to the structural complexity of their internal representations (excluding input and output layers), and this rank ordering corresponds both to classes of psychological processes and to empirical criteria. Two-layered nets have no internal representation. Three-layered nets contain a representation that is computed from the input. While allowing that there are many variations, and potential for development, the representation in a typical three-layered net is "holistic" and is not structured in a way that meets the criteria for representation of relations. Three-layered nets can represent content-specific information and can form prototypes (Quinn & Johnson, 1997) but they lack compositionality and systematicity (Fodor & Pylyshyn, 1988; Phillips, 1994). They can only mediate transfer based on similar content (Marcus, submitted) and not between isomorphic structures with different contents (Phillips & Halford, 1997).

Nets that model higher cognitive processes should implement the properties of relational processes defined above. There are currently a number of competing models that can meet these criteria, discussed by (Halford et al., in press). In the model we will present here, each relational instance is represented as a unique n-tuple, by representing bindings between relation symbol and arguments as outer products. Thus to represent loves(Joe,Jenny), each component, *loves*, *Joe* and *Jenny* is represented as a vector, and the binding is represented as the outer product of these vectors. The outer product corresponds to the binding units, shown for Rank 2 in Figure 1 but omitted for simplicity at higher ranks. Other instances of *loves* are represented in the same way, and can be summed to form a tensor product which represents the relation *loves* (Halford et al., in press, section 4.1.1.2). Thus loves(Joe,Jenny) and loves(Tom,Wendy) are represented as:

$$V_{loves} \oplus V_{Joe} \oplus V_{Jenny} + V_{loves} \oplus V_{Tom} \oplus V_{Wendy}$$

Neural net representations of relations from unary to quinary are shown schematically in the rightmost column of Figure 2. An n-ary relation is represented by the rank-n tensor, $V_R \oplus V_{a1} \oplus, \ldots, \oplus V_{an}$. A unary relation such as happy(John) is represented by the outer product of vectors representing "happy" and "John": $V_{happy} \oplus V_{John}$. In Figure 2 the two vectors are bound by a set of connections to a matrix of binding units. Rank 2 is the lowest structural level, but the transition from Rank 1 to Rank 2 can be envisaged by imagining the hidden layer at Rank 1 (Figure 2) being divided into two components which are then connected so as to form a matrix as shown for Rank 2. More complex relations are represented by tensor products of higher rank. A binary relation is represented by $V_R \oplus V_{a1} \oplus V_{a2}$, and so on. There is one component representing the symbol and one for each argument, so the representation of an n-ary relation has n+1 components. The components retain their identity, and the representations have the compositionality property. The model provides a natural explanation for empirical observations that cognitive processing load increases with relational complexity (Halford et al., in press, Section 5.). Representation of a relation of rank r with m units in each vector, requires m^r bindings units. The model implements all properties of relational knowledge (Halford et al., in press, Section 4.2) and is more efficient than models based on role-filler bindings for data bases in which relational instances are superimposed in the sense that role-filler bindings require r units per relational instance, where symbol-argument bindings require 1 unit per instance (Halford et al., in press, sections 2.2.1.2 and 4.1.3).

ASSOCIATIONS, RELATIONS AND ANALOGY

It follows from this analysis that higher cognitive processes differ from associative processes in that the former entail representation and processing of structure. A task is cognitive to the extent that it entails a representation and processing of the structure of the task or situation. The representation should have the properties identified above. Representation of structure (relations) is essential to analogy, and this principle can be used to devise what is probably the most objective and straightforward test for cognitive processes.

The essential idea is that if the structure of a task is learned, it can be transferred to isomorphs using analogical mapping, and unknown items in the new task can be predicted. This principle has been applied successfully with tasks based on mathematical groups (Halford, Bain, Maybery, & Andrews, in press) but can be easily illustrated with the conditional discrimination task summarised in Table 2. Suppose someone has learned the original task. While this can be done by configural association, as noted above, configural discrimination does not lead to a representation of structure because the elements lose their identity. However the task can also be learned by acquiring a representation of structure. The two modes of learning can be distinguished because only representation of structure enables transfer to isomorphs with prediction of new items. Notice that, in the transfer task in Table 2, once the first item is known and is mapped into the structure, the remaining three items can be easily predicted, irrespective of order of presentation.

Prediction of unknown items in an isomorphic task in this way requires analogical mapping, which in turn requires representation of structure. It is not possible if the task has been learned by configural association. Therefore transfer between isomorphs, with prediction of unseen items is a clearcut and objective measure of structural processing. It is a good way to assess higher cognitive processes. Notice too that it does not impose any extraneous task demands. The isomorphic task is assessed by the same procedure as the original task, and structure processing can be assessed by the number of correct items on the first trial of a new problem. It is not necessary to ask participants to describe the structure or to define rules, both of which impose an additional demand for articulation. We have been able to use this methodology successfully (Halford, 1980; Halford, Bain, et al., in press; Halford & Wilson, 1980) and have found that was related in a systematic way to other criteria.

CATEGORIES, STRUCTURE AND SIMILARITY

Although natural categories can be based on prototypes, prototypes do not represent structure (Medin, 1989). This problem can be overcome by forming prototypes based on relational instances. Relational instances such as Lives-in (chair, living room), Lives-in (vase, living room), Lives-in(couch, living room) can be represented as outer products of vectors and superimposed on a tensor product. The superimposed representation automatically averages features of the relational instances and corresponds to a prototype of living room furniture, but it also incorporates structure in the form of propositional information.

Similarity depends on more than common features, and is influenced by structure. For example grey hair is rated more similar to white hair than to black hair, whereas grey clouds are more similar to black clouds than to white clouds, because of our intuitive theories of ageing and weather respectively (Medin, 1989). Our model can handle similarity based both on elements and structure.

Element similarity can be assessed by computing the dot (inner) products of vectors representing two elements. If "desk", "chair" and "vase" were coded by vectors representing sets of features, the dot products of vectors representing "desk" and "chair" would be higher than dot products of vectors representing "desk" and "vase", reflecting greater similarity in the former pair.

Structural similarity can be handled by computing dot products of tensor products. The propositions feeds(soup-kitchen,woman) and feeds(woman,squirrel) have low similarity because "woman" occupies different roles. If we represent the propositions respectively as: $v_{feeds} \oplus v_{soup\text{-}kitchen} \oplus v_{woman}$ and $v_{feeds} \oplus v_{woman} \oplus v_{squirrel}$, the dot products of these tensors will have a low value (expected value is zero with orthonormal vectors, low with sparse random vectors). This reflects the relational context, because woman is bound to soup-kitchen in one case and squirrel in the other. However cases such as feeds(man,woman) and feeds(woman,man) are distinguished solely by the roles occupied by entities "woman" and "man". We represent these in analogous fashion as $v_{feeds} \oplus v_{man} \oplus v_{woman}$ and $v_{feeds} \oplus v_{woman} \oplus v_{man}$. Dot products of these vectors will again be low, reflecting "man" and "woman" being in different roles. This occurs because dot products are computed so as to respect structural alignment (the elements of v_{man} are multiplied by the elements of v_{woman}, and vice-verse, giving the dot product an expected value of zero with orthonormal vectors, or a low value with sparse random vectors). This illustrates the sensitivity of the model to structural alignment.

RELATIONAL CONTEXT SIMILARITY

The similarity of two items can be based on the degree to which they are used in the same relational context. For example, in the relational domain constructed around the items chair, desk and vase detailed above, chair and desk would achieve a high similarity as they both occur frequently in the same relational context (ie. Made_of(chair, wood) and Made_of(desk, wood), Stands_on(chair, floor) and Stands_on(desk, floor)). Chair and vase, however would achieve a lower similarity as they occur less frequently in the same relational context. Furthermore "woman" in feeds(soup-kitchen,woman) is dissimilar to "woman" in feeds(woman,squirrel) because the relational contexts are different, "soup-kitchen" in one case and "squirrel" in the other.

The relational context similarity of two items, *a* and *b* is computed as a normalised dot product of the rank 2 tensors retrieved from computing the dot product of each item's vector against an appropriate dimension of the rank 3 tensor storing the relations[1]. This can be applied to the hair-colour and cloud-colour examples above. We will represent a naive theory of ageing by propositions such as old-people-have(hair,grey),old-people-have(hair,white),young-people-have(hair,black),young-people-have(hair,brown) etc. These propositions can be superimposed on a tensor product representation. If we query this representation with "grey" we retrieve "old-people-have(hair,_)". If we query it with "white" we retrieve "old-people-have(hair,_)". The dot products of these tensors will be high, reflecting high similarity. However if we query the representation with "black" we retrieve "young-people-have(hair,_) and the dot product of this with "old-people-have(hair,_)" is low.

By contrast, our knowledge of weather is represented by propositions threatening(clouds,grey), threatening(clouds,black), nonthreatening(clouds,white) etc. Querying with "grey" and "black" yields threatening(clouds,_) in both cases, with high dot products representing high similarity. Querying with "white" yields nonthreatening(clouds,_) which is dissimilar to threatening(clouds,_). Thus the model represents naive theories as sets of propositions coded in a tensor product. Relational context, as defined above, accounts for the effect of naive theories on similarity.

Representational ranks are really points on a continuum, and limits on processing capacity are soft, so performance declines gracefully as the rank demanded by a task increases. It is proposed to model performances of intermediate rank, using the graceful degradation and graceful saturation properties of tensor products (Wilson & Halford, 1994).

EMPIRICAL INDICATORS OF RANKS

Each rank has a unique set of empirical indicators. We will consider the main indicators for each rank.

Rank 0 is indicated by elemental association. Since this is evidently universal to all animals with nervous systems, no special predictions are made.

Rank 1 is best assessed by conditional discrimination. It is indicated in general by tasks that require content-specific representations. Representation of vanished objects and prototype formation both entail this requirement, and are performed by infants 3-6 months (Baillargeon, 1987). Consequently the theory predicts that with suitable testing and training techniques, infants of this age can acquire conditional discrimination. The significance of this can be seen from the fact that in the past children under five years have had great difficulty with this task (Rudy, 1991). Two further predictions follow. The first is that transfer to isomorphs of conditional discrimination will not be possible until a median age of five years. The second is that formation acquisition of conditional discrimination will be related to representation of vanished objects as assessed by Baillargeon (1987) and to prototype formation.

Rank 2 entails a relation-symbol that is independent of the entity to which it is bound, and is the simplest symbolic representation. Tasks that require this level of structure include:

Explicit category membership, such as dog(Rover), where the category label *dog* is represented independently of the entity to which it is bound, *Rover*. As with all relations, the argument slot functions as a variable, and can be instantiated in a variety of ways such as dog(Fido), dog(Penny) etc. Representation of explicit categories, in which there is a binding between a category symbol and instances of the category, seems to occur at approximately one year (Gershkoff-Stowe, Thal, Smith, & Namy, 1997; Sugarman, 1982).

Inferences about numerosity based on category membership Xu and Carey (1996).

Word comprehension, or understanding that words function as symbols for their referents.

Representing the binding between an object and its location, as assessed in the A-not B task (Halford, 1993, pp. 51-56; Wellman, Cross, & Bartsch, 1986).

Match-to-sample requires choosing an object that matches the sample (e.g. if shown an apple as sample, required to choose between an apple and a hammer). This task has been analysed by Premack (1983) and Halford et al. (in press) and is an analogy based on a unary relation. Transfer to an isomorphic task (e.g. the sample is a hammer, and the choices are a banana and a hammer) demonstrates the principle is recognised independently of specific content.

This theory appears to be unique in predicting a correspondence between all five tasks.

Rank 3 entails symbolic processes based on binary relations, which develop at a median age of two years (Halford, 1993). Tasks that can be used to test this level of performance include:

Binary relational match-to-sample requires choice of a pair of objects that has the same relation as the sample (e.g. if the sample is XX, they should choose AA rather than BC. If the sample is XY, they should choose BC rather than AA). This implies a form of analogical reasoning based on binary relations, a Rank 3 representation (Gentner & Stevens, 1983; Halford et al., in press; Holyoak & Thagard, 1995).

Sorting into two categories can be assessed using the technique of Gershkoff-Stowe et al. (1997. Balance scale - weight and distance rules requires children to decide whether a beam should balance, or which side will go down, based on either weight or distance, with the other factor held constant [Halford, 1995 #2927). This requires binary relations (Halford et al., in press, Section 6.3.1).

Rank 4 entails ternary relations. This level of structure is required for transitive inference, hierarchical classification, class inclusion, hypothesis testing, cardinality and comprehension of sentences (Andrews, 1996; Halford, 1993; Halford et al., in press). Other tests that require this level of structure include:

Transfer between isomorphs of conditional discrimination tasks with prediction of unseen items. Conditional discrimination has a well defined structure that can be assessed by transfer to isomorphs. As pointed out above, if the relations in the original task in Table 1 are learned, and given any one item of the isomorphic transfer task, the remaining three items can be predicted, irrespective of order of presentation. This is a case of analogical reasoning (Gentner & Stevens, 1983; Holyoak & Thagard, 1995) in which the structure of the original task (the base or source) is mapped into the transfer task (target). The structure of conditional discrimination is basically a ternary relation, in that it consists of ordered 3-tuples (e.g. colour,shape,response). Therefore, while original learning can be used to infer nothing more complex than configural association (*Rank 1*), prediction of unseen items of a new isomorphic transfer task reflects processing ternary relations (*Rank 4*). The same paradigm can be used to assess two different levels of cognitive process, with procedure held constant and without additional demands such as articulation. Infants should be able to learn the original discrimination but should not be able to predict unseen items on the isomorphic transfer task. Five year olds should be able to do both. These predictions are more optimistic than previous findings that conditional discrimination is not learned before age 5 (Gollin, 1966; Rudy, 1991).

The tendency to prefer reversal over nonreversal shifts (Kendler, 1995). Ability to make efficient reversal shifts in multidimensional discrimination problems was first analysed in detail by Kendler and Kendler (1962) and there is a long history of research (see review by Kendler, 1995, and commentary by Halford, 1997). Reversal shifts depend on representation of the relevant dimension, which requires processing a ternary relation, because a dimension is a set on which an asymmetric, transitive relation is defined. Representation of a dimension requires induction of a relational schema (Halford, Bain, et al., in press). Consequently this longstanding enigma can be explained as a form of relational processing. Many predictions follow from this, but the one on which we will focus here is that preference of reversal shifts should correspond to other ternary relations tasks.

Ranks 5 and 6 entail quaternary and quinary relations respectively. Rank 5 is typically understood at age 11 (Halford, 1993), but there is virtually no useable data on Rank 6, though it is believed to occur only in a minority of adults. However we will consider two tasks that appear to require this level of processing, but have not been analysed in this way before.

RELATIONAL PROCESSES IN REASONING

In this section we will consider how relational processes could be involved in two reasoning tasks, knights and knaves and categorical syllogisms.

Knights and knaves problems are based on the following scenario. Suppose there is an island where there are just two sorts of inhabitants - *knights* who always tell the truth and *knaves* who always lie. An example problem is: *A says "I am a knave and B is a knave". B* *says, "A is a knave". What is the status of A and B: Knight, knave, or impossible to tell?* (Rips, 1989, pp. 85-86). The solution entails two or more steps, but we focus on the step that requires the highest relational complexity: If we assume *A* is a knight, then *A*'s statement that *A* and *B* are knaves must be true, but *A* says *A* is a knave, which is a contradiction. Therefore *A* must be a knave. Symbolically:

kt($\underline{A}$) and says($\underline{A}$,(kv($\underline{A}$) and kv($\underline{B}$))) Æ kv($\underline{A}$).

Using the type of analysis developed by Halford et al. (in press-a) there are five variables in this expression, corresponding to the five underlined arguments. Therefore this inference is quinary. The second step is to reason that if it is false that *A* and *B* are knaves, and that *A* is a knave, then *B* must be a knight: false(kv($\underline{A}$) and kv($\underline{B}$)) and kv($\underline{A}$) Æ kt($\underline{B}$). This step is quaternary, so task complexity, defined by the most complex step, is quinary.

Categorical syllogism tasks have been more extensively investigated, but we will focus on the following example tasks:

All *A* are *B*, all *B* are *C*. This would be represented by Johnson-Laird & Byrne (1991, Table 6.1) as the mental model: [[a]b]c. This mental model can be expressed as a relation between the following classes of entities (where ¬*A* means "not *A*"): *ABC*, ¬*ABC*, ¬*A*¬*BC*. We can think of this as follows: There is one class of entities with properties *A*,*B* and *C*, another class with properties not-*A*, *B* and C, and another class with properties not -A, not -B and C. The mental model that relates these three classes has the complexity of a ternary relation. Now consider the syllogism:

Some *A* are *B*, No *B* are *C*. The premises express a relation between the following classes of entities: *A*¬*BC*, *A*¬*B*¬*C*, *AB*¬*C*, ¬*A*¬*BC*, ¬*AB*¬*C* (c.f. J-L&B, 1991, Table 6.1). The problem relates 5 classes of entities, so it has the complexity of a quinary relation. J-L&B define complexity in terms of the number of mental models required for a problem. The first problem above requires one model and is easy (88% correct) while the second requires 3 models and is difficult (38% correct). However more difficult problems tend to entail more complex relations. Of the 27 syllogisms with valid conclusions, there are 7 with ternary relations that entail 1 mental model, and 17 with relations more complex than ternary that entail more than 1 mental model (contingency coefficient C = .61). Therefore the relational complexity metric has potential to provide an alternative explanation to number of mental models for difficulty of categorical syllogisms.

CONCLUSION

We wish to propose that the representation and processing of structure, including analogical mapping, are core processes in higher cognition. They can be used as criteria for distinguishing tasks that demand higher cognitive processes from those that can be performed by more basic processes. Ability to form analogies can also be used as criterion for neural net models of higher cognitive processes. The relational complexity metric permits levels of structure to be distinguished.

Cognitive tasks can be grouped into equivalence classes of equal structural complexity, and the classes can be ordered according to representational rank. Ranks 0 and 1 are associative, do not entail explicit representation of structure, and do not enable analogical mappings to be made. Ranks 2-6 entail explicit representation of relations, from unary to quinary. In general they have the properties normally attributed to higher cognitive processes. There is a correspondence between three domains: level of structural complexity, neural net architecture, and observable properties of performance.

REFERENCES

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Andrews, G. (1996, August 1996). Assessment of relational reasoning in children aged 4 to 8 years. Paper presented at the Conference of the International Society for Study of Behavioral Development, Quebec City

Baillargeon, R. (1987). Young infants' reasoning about the physical and spatial properties of a hidden object. Cognitive Development, 2, 179-200.

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. Psychological Review, 85, 1-21.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. Cognitive Psychology, 17, 391-416.

Chomsky, N. (1980). Rules and representations. The Behavioral and Brain Sciences, 3, 1-61.

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. Mind and Language, 8(4), 487-519.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. Artificial Intelligence, 41, 1-63.

Fodor, J. A. (1975). The language of thought. New York: Thomas Y. Crowell Company.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. Cognition, 28, 3-71.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7, 155-170.

Gentner, D., & Stevens, A. L. (1983). Mental models. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gershkoff-Stowe, L., Thal, D. J., Smith, L. B., & Namy, L. L. (1997). Categorization and its developmental relation to early language. Child Development, 68, 843-859.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15, 1-38.

Gollin, E. S. (1966). Solution of conditional discrimination problems by young children. Journal of Comparative and Physiological Psychology, 62, 454-456.

Gray, B., Halford, G. S., Wilson, W. H., & Phillips, S. (1997, September 26-28). A Neural Net Model for Mapping Hierarchically Structured Analogs. Proceedings of the Fourth conference of the Australasian Cognitive Science Society, University of Newcastle.

Halford, G. S. (1980). A learning set approach to multiple classification: Evidence for a theory of cognitive levels. International Journal of Behavioral Development, 3, 409-422.

Halford, G. S. (1993). Children's understanding: the development of mental models. Hillsdale, NJ: Erlbaum.

Halford, G. S. (1997). Review of Levels of Cognitive Development. Tracey S. Kendler. Mahwah, NJ: Erlbaum, 1995. Pp vii + 187. Merrill-Palmer Quarterly, 43(4), 694-699.

Halford, G. S., Bain, J. D., Maybery, M. T., & Andrews, G. (in press). Induction of relational schemas: Common processes in reasoning and learning set acquisition. Cognitive Psychology.

Halford, G. S., Wiles, J., Humphreys, M. S., & Wilson, W. H. (Eds.). (1993). Parallel distributed processing approaches to creative reasoning: Tensor models of memory and analogy. Menlo Park, CA: AAAI Press.

Halford, G. S., & Wilson, W. H. (1980). A category theory approach to cognitive development. Cognitive Psychology, 12, 356-411

Halford, G. S., Wilson, W. H., Guo, J., Gayler, R. W., Wiles, J., & Stewart, J. E. M. (1994). Connectionist implications for processing capacity limitations in analogies. In K. J. Holyoak & J. Barnden (Eds.), Advances in connnectionist and neural computation theory, Vol. 2: Analogical connections (pp. 363-415). Norwood, NJ: Ablex.

Halford, G. S., Wilson, W. H., & Phillips, S. (in press). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. Behaviorial and

Brain Sciences.

Holyoak, K. J., & Thagard, P. (1995). Mental leaps. Cambridge, MA: MIT Press.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. Psychological Review, 104, 427-466.

Humphrey, G. (1951). Thinking: An introduction to its experimental psychology. London: Methuen.

Hunt, E. B. (1962). Concept learning: An information processing problem. New York: Wiley.

Johnson-Laird, P.N. (1983). Mental models. Cambridge: Cambridge University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). Deduction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80(4), 237-251.

Kendler, H. H., & Kendler, T. S. (1962). Vertical and horizontal processes in problem solving. Psychological Review, 69, 1-16.

Kendler, T. S. (1995). Levels of cognitive development. Mahwah, NJ: Erlbaum.

Kohler. (1957). The Mentality of Apes (2nd rev. ed. Ella Winter, Trans). Harmondsworth: Penguin.

Lassaline, M. (1996). Structural alignment in induction and similarity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(3), 754-770.

Marcus, G. F. (submitted). Rethinking eliminative connectionism.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. Cognitive Psychology, 25(4), 431-467.

Medin, D. L. (1989). Concepts and conceptual structure. American Psychologist, 44(12), 1469-1481.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). Plans and the structure of behavior. New York: Holt, Rinehart and Winston.

Minsky, M., & Papert, S. (1969). Perceptrons: An introduction to computational geometry. Cambridge, MA.: MIT Press.

Mitchell, M., & Hofstadter, D. R. (1990). The emergence of understanding in a computer model of concepts and analogy-making. Physica D, 42(1-3), 322-334.

Neal, A., & Hesketh, B. (1997). Future directions for implicit learning: Toward a clarification of issues associated with knowledge representation and consciousness. Psychonomic Bulletin and Review, 4(1), 73-78.

Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

Niklasson, L., & van Gelder, T. (1994,). Can connectionist models exhibit non-classical structure sensitivity? Sixteenth Annual Conference of the Cognitivie Science Society, Atlanta, Georgia, 664-669.

Phillips, S. (1994). Strong systematicity within connectionism: The tensor-recurrent network. Sixteenth Annual Conference of the Cognitive Science Society, 723-727.

Phillips, S., & Halford, G. S. (1997,). Systematicity: Psychological evidence with connectionist implications. Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society, Stanford University, 614-619.

Phillips, S., Halford, G. S., & Wilson, W. H. (1995, July). The processing of associations versus the processing of relations and symbols: A systematic comparison. Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, Pittsburgh, PA, 688-691.

Piaget, J. (1950). The psychology of intelligence. (M. Piercy & D. E. Berlyne, Trans.) London: Routledge & Kegan Paul, (Original work published 1947).

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. Pychological Review, 102, 533-566.

Premack, D. (1983). The codes of man and beasts. The Behavioral and Brain Sciences, 6, 125-167.

Quinn, P. C., & Johnson, M. H. (1997). The emergence of perceptual category representations in young infants: A connectionist analysis. Journal of Experimental Child Psychology, 66, 236-263.

Rips, L. J. (1989). The psychology of knights and knaves. Cognition, 31, 85-116.

Rudy, J. W. (1991). Elemental and configural associations, the hippocampus and development. Developmental Psychobiology, 24(4), 221-236.

Schmajuk, N. A., & DiCarlo, J. J. (1992). Stimulus configuration, classical conditioning, and hippocampal function. Psychological Review, 99, 268-305.

Siegler, R. S. (1981). Developmental sequences within and between concepts. Monographs of the Society for Research in Child Development, 46, 1-84.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119, 3-22.

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. Cognitive Science, 16, 1-40.

Smolensky, P. (1988). On the proper treatment of connectionism. Behavioral and Brain Sciences, 11(1), 1-74.

Spearman, C. E. (1923). The nature of intelligence and the principles of cognition. London: MacMillan.

Sugarman, S. (1982). Developmental change in early representational intelligence: Evidence from spatial classification strategies and related verbal expressions. Cognitive Psychology, 14, 410-449.

Tomasello, M., & Call, J. (1997). Primate cognition. New York: Oxford University Press.

Van Gelder, T., & Niklasson, L. (1994,). Classicalism and cognitive architecture. Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society, Atlanta, Georgia, 905-909.

Wellman, H. M., Cross, D., & Bartsch, K. (1986). Infant search and object permanence: A meta-analysis of the A-not-B error. Monographs of the Society for Research in Child Development, 51.

Wertheimer, M. (1945). Productive thinking. New York: Harper.

Wilson, W. H., & Halford, G. S. (1994,). Robustness of tensor product networks using distributed representations. Proceedings of the Fifth Australian Conference on Neural Networks, Brisbane, 258-261.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. Cognitive Psychology, 30, 111-153.

Emotional Analogies and Analogical Inference

Paul Thagard
Philosophy Department
University of Waterloo
Waterloo, Ontario, N2L 3G1

1. INTRODUCTION

Despite the growing appreciation of the relevance of affect to cognition, analogy researchers have paid remarkably little attention to emotion. This paper discusses three general classes of analogy that involve emotions. The most straightforward are analogies and metaphors *about* emotions, for example "Love is a rose and you better not pick it." Much more interesting are analogies that involve the transfer of emotions, for example in empathy in which people understand the emotions of others by imagining their own emotional reactions in similar situations. Finally, there are analogies that generate emotions, for example analogical jokes that generate emotions such as surprise and amusement.

Understanding emotional analogies requires a more complex theory of analogical inference than has been currently available, and section 2 presents a new account that shows how analogical inference can be defeasible, holistic, multiple, and emotional, in ways to be described. Analogies about emotions can to some extent be explained using the standard models such as ACME and SME, but analogies that transfer emotions require an extended treatment that appreciates the special character of emotional states. I describe HOTCO, a new model of emotional coherence, that simulates transfer of emotions. Finally, I show how HOTCO models the generation of emotions such as reactions to humorous analogies.

2. ANALOGICAL INFERENCE: CURRENT MODELS

In logic books, analogical inference is usually presented by a schema such as the following (Salmon, 1984, p. 105):

Objects of type *X* have properties *G*, *H*, etc.
Objects of type *Y* have properties *G*, *H*, etc.
Objects of type *X* have property *F*.
Therefore: Objects of type *Y* have property *F*.

For example, when experiments determined that large quantities of the artificial sweetener saccharine caused bladder cancer in rats, scientists analogized that it might also be carcinogenic in humans. Logicians routinely point out that analogical arguments may be strong or week depending on the extent to which the properties in the premises are relevant to the property in the conclusion.

This characterization of analogical inference, which dates back at least to John Stuart Mill's nineteenth-century *System of Logic*, is flawed in several respects. First, logicians rarely spell out what "relevant" means, so the schema provides little help in distinguishing strong analogies from weak. Second, the schema is stated in terms of objects and their properties, obscuring the fact that the strongest and most useful analogies involve relations, in particular causal relations (Gentner, 1983; Holyoak and Thagard, 1995). Such causal relations are usually the key to determining relevance: if, in the above schema, *G* and *H* together cause *F* in *X*, then analogically they may cause *F* in *Y*, producing a much stronger inference than just counting properties. Third, logicians typically discuss analogical arguments and tend to ignore the complexity of analogical inference, which requires a more holistic assessment of a potential conclusion with respect to other information. There is no point in inferring that objects of type *Y* have property *F* if you already know of many such objects that lack *F*, or if a different analogy suggests that they do not have *F*. Analogical inference must be defeasible, in that the potential conclusion can be overturned by other information, and it must be holistic in that everything the inference maker knows is potentially relevant to overturning or enhancing the inference.

Compared to the logician's schema, much richer accounts of the structure of analogies have been provided by computational models of analogical mapping such as SME (Falkenhainer, Forbus, and Gentner, 1989) and ACME (Holyoak and Thagard, 1989). SME uses relational structure to generate candidate inferences, and ACME transfers information from a source analog to a target analog using a process that Holyoak, Novick and Melz (1994) called copying with substitution and generation (CWSG). Similar processes are used in case-based reasoning (Kolodner, 1993), and in many other computational models of analogy.

But all of these computational models are inadequate for understanding analogical inference in general and emotional analogies in particular. They do not show how analogical inference can be defeasible, holistic, and multiple – making use of more than one analogy to support or defeat a conclusion. Moreover, the prevalent models of analogy encode information symbolically and assume that what is inferred is verbal information that can be represented in propositional form by predicate calculus or some similar representational system.[1] But as section 5 documents, analogical inference often serves to transfer an emotion, not just the verbal representation of an emotion. I will now describe how a new model of emotional coherence, HOTCO, can perform analogical inferences that are defeasible, holistic, multiple, and emotional.

3. ANALOGICAL INFERENCE IN HOTCO

I recently proposed a theory of emotional coherence that has applications to numerous important psychological phenomena such as trust (Thagard, forthcoming). This theory makes the following assumptions about inference and emotions:

1) All inference is coherence-based. So-called rules of inference such as *modus ponens* do not by themselves license inferences, because their conclusions may contradict other accepted information. The only rule of inference is: Accept a conclusion if its acceptance maximizes coherence.
2) Coherence is a matter of constraint satisfaction, and can be computed by connectionist and other algorithms (Thagard and Verbeurgt, 1998).
3) There are six kinds of coherence: analogical, conceptual, explanatory, deductive, perceptual, and deliberative (Thagard, Eliasmith, Rusnock, and Shelley, forthcoming).
4) Coherence is not just a matter of accepting or rejecting a conclusion, but can also involve attaching a positive or negative emotional assessment to a proposition, object, concept, or other representation.

From this coherentist perspective, inference takes on a very different complexion from what is suggested by logical deduction. Philosophers who have advocated coherentist accounts of inference include Bosanquet (1920) and Harman (1986).

The computational model HOTCO (for "hot coherence") implements these theoretical assumptions. It amalgamates the following previous coherence models of coherence:

- Explanatory coherence: ECHO (Thagard, 1989, 1992);

---

[1] One of the few attempts to deal with nonverbal analogies is the VAMP system for visual analogical mapping: Thagard, Gochfeld, and Hardy (1992).

- Conceptual coherence: IMP (Kunda and Thagard, 1996);
- Analogical coherence: ACME (Holyoak and Thagard, 1989);
- Deliberative coherence: DECO (Thagard and Millgram, 1995).

Amalgamation is natural, because all of these models use a similar connectionist algorithm for maximizing constraint satisfaction, although they employ different constraints operating on different kinds of representation. What is novel about HOTCO is that representational elements possess not only activations that represent their acceptance and rejection, but also valences that represent a judgment of their positive or negative emotional appeal. In HOTCO, as in its component models, inferences about what to accept are made by a holistic process in which activation spreads through a network of units with excitatory and inhibitory links, representing elements with positive and negative constraints. But HOTCO spreads valences as well as activations in a similar holistic fashion, using the same system of excitatory and inhibitory links. For example, HOTCO models the decision of whether to hire a particular person as a babysitter as in part a matter of "cold" deliberative, explanatory, conceptual, and analogical coherence, but also as a matter of generating an emotional reaction to the candidate. The emotional reaction derives from a combination of the cold inferences made about the person and the valences attached to what is inferred. For example, if you infer that that a babysitting candidate is responsible, intelligent, and likes children, the positive valence of these attributes will spread to him or her; whereas if coherence leads to you infer that the candidate is lazy, dumb, and psychopathic, he or she will acquire a negative valence. In HOTCO, valences spread through the constraint network in much the same way that activation does (see Thagard, forthcoming, for technical details).

Now I can describe how HOTCO performs analogical inference in a way that is defeasible, holistic, multiple, and emotional. HOTCO uses ACME to perform analogical mapping between a source and a target, and copying with substitution and generation to produce new propositions to be inferred. It can operate either in a broad mode in which everything about the source is transferred to the target, or in a more specific mode in which a query is used to enhance the target using a particular proposition in the source. Here, in predicate calculus formalization where each proposition has the structure (predicate (objects) proposition-name), is an example of scientific inference (Shelley forthcoming):

Source 1: centroscymnus

(have (centroscymnus rod-pigment-1) have-1

(absorb (rod-pigment-1 472nm-light) absorb-1)

(penetrate (472nm-light deep-ocean-water) penetrate-1)

(see-in (centroscymnus deep-ocean-water) see-in-1)

(inhabit (centroscymnus deep-ocean-water) inhabit-1)

(enable (have-1 see-in-1) enable-1)

(because (absorb-1 penetrate-1) because-1)

(adapt (see-in-1 inhabit-1) adapt-1)

Target: coelacanth-3

(have (coelacanth rod-pigment-3) have-3)

(absorb (rod-pigment-3 473nm-light) absorb-3)

(penetrate (473nm-light deep-ocean-water) penetrate-3)

(see-in (coelacanth deep-ocean-water) see-in-3)

(enable (have-3 see-in-3) enable-3)

(because (absorb-3 penetrate-3) because-3)

Operating in specific mode, HOTCO is asked what depth the coelacanth inhabits, and uses the proposition INHABIT-1 in the source to construct for the target the proposition (inhabit (coelacanth deep-ocean-water) inhabit-new)

Operating in broad mode and doing general CWSG, HOTCO can analogically transfer everything about the source to the target, in this case generating the same proposition as a candidate to be inferred.

However, HOTCO does *not* actually infer the new proposition, because analogical inference is defeasible. Rather, it simply establishes an excitatory link between the unit representing the source proposition INHABIT-1 and the target proposition INHABIT-NEW. This link represents a positive constraint between the two propositions, so that coherence maximization will encourage them to be accepted together or rejected together. The source proposition INHABIT-1 is presumably accepted, so in the HOTCO model it will have positive activation which will spread to provide positive activation to INHABIT-NEW, unless INHABIT-NEW is incompatible with other accepted propositions that will tend to suppress its activation. Thus analogical inference is defeasible, because all HOTCO does is to create a link representing a new constraint for overall coherence judgment, and it is holistic, because the entire constraint network can potentially contribute to the final acceptance or rejection of the inferred proposition.

Within this framework, it is easy to see how analogical inference can employ multiple analogies, because more than one source can be used to create new constraints. Shelley (forthcoming) describes how biologists do not simply use the centroscymnus analog as a source to infer that coelacanths inhabit deep water, but also use the following different source:

Source 2: ruvettus-2

(have (ruvettus rod-pigment-2) have-2)

(absorb (rod-pigment-2 474nm-light) absorb-2)

(penetrate (474nm-light deep-ocean-water) penetrate-2)

(see-in (ruvettus deep-ocean-water) see-in-2)

(inhabit (ruvettus deep-ocean-water) inhabit-2)

(enable (have-2 see-in-2) enable-2)

(because (absorb-2 penetrate-2) because-2)

(adapt (see-in-2 inhabit-2) adapt-2)

The overall inference is that coelacanths inhabit deep water because they are like the centroscysmus and the ruvettus sources in having rod pigments that are an adaptation to deep water. Notice that these are deep, systematic analogies, because the theory of natural selection suggests that the two source fishes have the rod pigments because they are adaptive for their deep ocean water environments. When HOTCO maps the ruvettus source to the coelecanth target after mapping the centroscysmus source, it creates links excitatory from the inferred proposition INHABIT-NEW with both INHABIT-1 in the first source and INHABIT-2 in the second source. Hence activation can flow from both these propositions to INHABIT-NEW, so that the inference is supported by multiple analogies. If another analog suggests a contradictory inference, then INHABIT-NEW will be both excited and inhibited. Thus multiple analogies can contribute to the defeasible and holistic character of analogical inference.

The new links created between the target proposition and the source proposition can also make possible emotional transfer. The coelacanth example is emotionally neutral, but if an emotional valence were attached to INHABIT-1 and INHABIT-2, then the excitatory links between them and INHABIT-NEW would make possible spread of that valence as well as spread of activation representing acceptance. Section 5 below provides detailed examples of this kind of emotional analogical inference.

4. ANALOGIES ABOUT EMOTIONS

The *Columbia Dictionary of Quotations* (available electronically as part of the Microsoft Bookshelf) contains many metaphors and analogies concerning love and other emotions. For example, love is compared to religion, a master, a pilgrimage, an angel/bird, gluttony, war, disease, drunkenness, insanity, market exchange, light, ghosts, and smoke. It is not surprising that writers discuss emotions non-literally, because it is very difficult to describe emotions straightforwardly in words. In analogies about emotions, verbal sources help to illuminate the emotional target, which may be verbally described but which also has an elusive, non-verbal, phenomenological aspect. Analogies are also used about negative emotions: anger is like a volcano, jealousy is a green-eyed monster, and so on.

In order to handle the complexities of emotion, poets often resort to multiple analogies, as in the following examples:

(1) John Donne:
Love was as subtly catched, as a disease;
But being got it is a treasure sweet.

(2) Robert Burns:
O, my love is like a red, red rose,
That's newly sprung in June:
My love is like a melodie,
That's sweetly play'd in tune.

(3) William Shakespeare:
Love is a smoke made with the fume of sighs,
Being purged, a fire sparkling in lovers' eyes,
Being vexed, a sea nourished with lovers' tears.
What is it else? A madness most discreet,
A choking gall and a preserving sweet.

In each of these examples, the poet uses more than one analogy or metaphor to bring out different aspects of love. The use of multiple analogies is different from the scientific example described in the last section, in which the point of using two marine sources was to support the same conclusion about the depths inhabited by coelacanths. In these poetic examples, different source analogs bring out different aspects of the target emotion, love.

Analogies about emotions may be general, as in the above examples about love, or particular, used to describe the emotional state of an individual. For example, in the movie *Marvin's Room,* the character played by Meryl Streep describes her reluctant to discuss her emotions by saying that her feelings are like fishhooks - you can't pick up just one. Just as it is hard to verbalize the general character of an emotion, it is often difficult to describe verbally one's own emotional state. Victims of post-traumatic stress disorder frequently use analogies and metaphors to describe their own situations (Meichenbaum (1994, pp. 112-113):

- I am a time bomb ticking, ready to explode.
- I feel like I am caught up in a tornado.
- I am a rabbit stuck in the glare of headlights who can't move.
- My life is like a rerun of a movie that won't stop.
- I feel like I'm in a cave and can't get out.
- Home is like a pressure cooker.
- I am a robot with no feelings.

In these particular emotional analogies, the target to be understood is the emotional state of an individual, and the verbal source describes roughly what the person feels like.

The purpose of analogies about emotions is often explanatory, describing the nature of a general emotion or a particular person's emotional state. But analogy can also be used to help deal with emotions, as in the following anonymous example:

Happiness is like a butterfly.
The more you chase it and chase it directly
the more it eludes you, but
if you sit quietly and turn your attention
to other things
it comes and softly sits on your shoulder.

People are also given advice on how to deal with negative emotions, being told for example to "vent" their anger, or to "put a lid on it."

In principle, analogies about emotions could be simulated by the standard models such as ACME and SME, with a verbal representation of the source being used to generate inferences about the emotional target. However, even in some of the above examples, the point of the analogy is not just to transfer verbal information, but also to transfer an emotional attitude. When someone says "I feel like I am caught up in a tornado," he or she may be saying something like "My feelings are like the feelings you would have if you were caught in a tornado." To handle the transfer of emotions, we need to go beyond verbal analogy.

5. ANALOGIES THAT TRANSFER EMOTIONS

As already mentioned, not all analogies are verbal: some involve transfer of visual representations (Holyoak and Thagard, 1995). In addition, analogies can involve transfer of emotions from a source to a target. There are at least three such kinds of emotional transfer, involved in persuasion, empathy, and self-explanation. In persuasion, I may use an analogy to convince you to adopt an emotional attitude. In empathy, I try to understand your emotional reaction to a situation by transferring to you my emotional reaction to a similar situation. In self-explanation, I try to get you to understand my emotion by comparing my situation and emotional response to it with situations and responses familiar to you.

The purpose of many persuasive analogies is to produce an emotional attitude, for example when at attempt is made to convince someone that abortion is abominable or that capital punishment is highly desirable. If I want to get someone to adopt positive emotions toward something, I can compare it to something else toward which he or she already has a positive attitude.. Conversely, I can try to produce a negative attitude by comparison with something already viewed negatively. The structure of persuasive emotional analogies is:

> You have an emotional appraisal of the source S.
>
> The target T is like S in relevant respects.
>
> So you should have a similar emotional appraisal of T.

Of course, the emotional appraisal could be represented verbally by terms such as "wonderful," "awful," and so on, but for persuasive purposes it is much more effective if the gut feeling that is attached to something can be transferred over to something else. For example, the point of analogizing using as sources such emotionally intense subjects as the Holocaust or infanticide is to transfer negative emotions to the target.

Blanchette and Dunbar (1997) thoroughly documented the use of persuasive analogies in a political context, the 1995 referendum in which the people of Quebec voted whether to separate from Canada. In three Montreal newspapers, they found a total of 234 different analogies, drawn from many diverse source domains: politics, sports, business, and so on. Many of these analogies were emotional: 66 were coded by Blanchette and Dunbar as emotionally negative, and 75 were judged to be emotionally positive. Thus more than half of the analogies used in the referendum had an identifiable emotional dimension. For example, the side opposed to Quebec separation said "It's like parents getting a divorce, and maybe the parent you don't like getting custody." Here the negative emotional connotation of divorce is transferred over to Quebec separation. In contrast, the yes side used positive emotional analogs for separation: "A win from the YES side would be like a magic wand for the economy."

HOTCO can naturally model the use of emotional persuasive analogies. The separation-divorce analogy can be represented as follows:

Source : divorce

(married (spouse-1 spouse-2) married-1)

(have (spouse-1 spouse-2 child) have-1)

(divorce (spouse-1 spouse-2) divorce-1) negative valence

(get-custody (spouse-1) get-custody-1)

(not-liked (spouse-1) get-custody-1) negative valence

Target: separation

(part-of (Quebec Canada) part-of-2)

(govern (Quebec Canada people-of-Quebec) govern-2) (separate-from (Quebec Canada) separate-from—2)

(control (Quebec people-of-Quebec) control-2)

When HOTCO performs a broad inference on this example (TO BE RUN), it should not only perform the analogical mapping from the source to the target and complete the target using copying with substitution and generation, but also transfer the negative valence attached to the proposition DIVORCE-1 to SEPARATE-FROM-2.

Persuasive analogies have been rampant in the recent debated about whether Microsoft has been engaging in monopolistic practices by including its World Wide Web browser in its operating system, Windows 98. In response to the suggestion that Microsoft also be required to include the rival browser produced by its competitor, Netscape. Microsoft's chairman Bill Gates complained that this would be "like requiring Coca-Cola to include three cans of Pepsi in very six-pack it sells," or like "ordering Ford to sell autos fitted with Chrysler engines." These analogies are in part emotional, since they are intended to transfer the emotional response to coercing Coca-Cola and Ford – assumed to be ridiculous – over to the coercion of Microsoft. On the other hand, critics of Microsoft's near-monopoly on personal computer operating systems have been comparing Gates to John D. Rockefeller, whose predatory Standard Oil monopoly on petroleum products was broken up by the U.S. government in 1911.

Another, more personal, kind of persuasive emotional analogy is identification, in which you identify with someone and then transfer positive emotional attitudes about yourself to them. According to Fenno (1978, p. 58), members of the U.S. congress try to convey a sense of identification to their constituents. The message is "You know me, and I'm like you, so you can trust me." The structure of this kind of identification is:

You have a positive emotional appraisal of yourself (source).

I (the target) am similar to you.

So you should have a positive emotional appraisal of me.

This is a kind of persuasive analogy, but differs from the general case in that the source and target are the people involved.

Empathy also involves transfer of emotional states between people; see Barnes and Thagard (1997) for a full discussion. It differs from persuasion in that the goal of the analogy is to understand rather than to convince someone. Summarizing, the basic structure is:

You are in situation T (target).

When I was in a similar situation S, I felt emotion E (source).

So maybe you are feeling an emotion similar to E.

As with persuasion and identification, such analogizing could be done purely verbally, but it is much more effective to actually feel something like what the target person is feeling. For example, if I want to understand the emotional state of a new graduate student just arrived from a foreign country, I can recall my emotional state of anxiety and confusion when I went to study in England. Here is a more detailed example of empathy involving someone trying to understand the distress of Shakespeare's Hamlet at losing his father by comparing it to his or her own loss of a job (from Barnes and Thagard, 1997):

Source: you	**Target: Hamlet**
fire (boss, you): s1-fire	kill (uncle, father): t1-kill
lose (you, job): s2-lose	lose (Hamlet, father): t2-lose
	marry (uncle, mother): t3-marry
cause (s1-fire, s2-lose): s3	cause (t1-kill, t2-lose): t3a

angry (you): s4-angry
depressed (you): s5-depressed
cause (s2-lose, s4-angry): s6
cause (s2-lose, s5-depressed): s7
indecisive (you): s8-indecisive
cause (s5-depressed, s8-indecisive): s9

angry (Hamlet): t4-angry
depressed (Hamlet): t5-depressed
cause (t2-lose, t4-angry): t6
cause (t2-lose, t5-depressed): t7

The purpose of this analogy is not simply to draw the obvious correspondences between the source and the target, but to transfer over your remembered image of depression to Hamlet.

Unlike persuasive analogies, whose main function is to transfer positive or negative valence, empathy requires transfer of the full range of emotional responses. Depending on his or her situation, I need to imagine someone being angry, fearful, disdainful, ecstatic, enraptured and so on. As currently implemented, HOTCO transfers only positive or negative valences associated with a proposition or object, but it can easily be expanded so that transfer involves an *emotional vector* which represents a pattern of activation of numerous units, each of whose activation represents different components of emotion. This expanded representation would also make possible the transfer of "mixed" emotions.

As an aside, let me speculate on the empathic origins of altruism. People are often altruistic, caring for the needs of others as well as for their own self-interests. From the perspective of evolutionary biology, altruism is a puzzle, because natural selection should favor behaviors that maximize the transmission of one's own genes, not those of others. Kin selection theory provides a plausible explanation for why social insects such as bees sacrifice themselves for their brothers and sisters, but barely begins to explain human altruism, which often extends beyond relatives. I conjecture that altruism is a byproduct of two other developments favored by natural selection: caring for relatives and analogy. First, it is plausible that genetic transmission is optimized by caring for one's children and for those also involved in caring for them. Such care is greatly aided by empathy - the ability to understand the emotional state of someone by analogy to one's own emotional state. But second, analogical inference is a general human capacity, not fully found in apes, but developing in children around the age of five (Holyoak and Thagard, 1995). Presumably, the ability to analogize was selected for as part of general selective pressures for increasing intelligence, although it may be that analogical inference is itself a byproduct of selection for other verbal and inferential abilities. It is even possible, I suppose, that analogical inference developed because it is socially valuable, for example in promoting empathy. In any case, assuming that both empathy for relatives and analogy developed biologically, altruism could have emerged as a byproduct. Our general analogical ability enables use to empathize with people in general, not just our immediate relatives, and thereby to attach value to the needs of others. Like the abilities to do mathematics, compose symphonies, philosophize, and play baseball, altruism was never directly selected for, but emerged as a byproduct of other valuable psychological capacities - empathy and analogy.

Empathy is only one kind of explanatory emotional analogy. In section 4, we already saw examples of analogies whose function is self-explanation, i.e. to explain one's own emotional state to another. The following news report describes an astronaut's emotional self-explanations:

> MOSCOW (December 2, 1997 1:53 p.m. EST Reuters) - Astronaut David Wolf says life on the Russian Mir space station can be distinctly unglamorous, with a load of chores that include cleaning the toilet and scrubbing fluff from air filters.
>
> Named NASA's inventor of the year in 1992, Wolf also describes feeling a wide array of emotions, including the moment when the U.S. space shuttle undocked from Mir, leaving him behind for his four-month mission.
>
> "I remember the place I last felt it. Ten years old, as my parents' station wagon pulled away from my first summer camp in southern Indiana. That satisfying thrill that something new is going to happen and we don't know what it is yet.
>
> "Life in space can also appear dream-like and cinematic, he said as he related being left in charge during another space walk, when he thought of Captain Kirk, hero of "Star Trek."
>
> "I felt like the kid in "Home Alone" as I assumed Tolya's usual posture at the central command post, the cockpit. Or, was it Kirk's position? Dream and reality run so close here."

Few people have the experience of being left in space, but most people can remember or imagine what it is like to leave for summer camp. Thus emotional analogies used for self-explanation have the function of enabling others to have an empathic understanding of oneself.

Here is a final example of analogical transfer of emotion: "Psychologists would rather use each other's toothbrushes than each other's terminology." This is complex, because at one level it is projecting the emotional reaction of disgust from use of toothbrushes to use of terminology, but it is also generating amusement. Let us now consider analogies that go beyond analogical transfer of emotions and actually generate new emotions.

6. ANALOGIES THAT GENERATE EMOTIONS

A third class of emotional analogies involves ones that are not about emotions and do not transfer emotional states, but rather serve to generate new emotional states. There are at least four subclasses of emotion-generating analogies, involving humor, irony, discovery, and motivation.

One of the most enjoyable uses of analogy is to make people laugh, generating the emotional state of mirth or amusement. The University of Michigan recently ran an informational campaign to get people to guard their computer passwords more carefully. Posters warn students to treat their computer passwords like underwear: make them long and mysterious, don't leave them lying around, and change them often. The point of the analogy is not to persuade anyone based on the similarity between passwords and underwear, but rather to generate amusement that focuses attention on the problem of password security.

A major part of what makes an analogy funny is a surprising combination of congruity and incongruity. Passwords do not fit semantically with underwear, so it is surprising when a good relational fit is presented (change them often). Other emotions can also feed into making an analogy funny, for example when the analogy is directed against a person or group one dislikes:

Why do psychologists prefer lawyers to rats for their experiments?

1. There are now more lawyers than rats;
2. The psychologists found they were getting

attached to the rats;

3. And there are some things that rats won't do.

This joke depends on a surprising analogical mapping between rats in psychological experiments and lawyers in their practices, and on negative emotions attached to lawyers. Another humorous analogy is implicit in the joke: "How can a single woman get a cockroach out of her kitchen? Ask him for a commitment."

Some analogical jokes depend on visual representations, as in the following children's joke: "What did the 0 say to the 8? Nice belt." This joke requires a surprising visual mapping between numerals and human dress. A more risqué visual example is. "Did you hear about the man with five penises? His pants fit like a glove." Here are a few more humorous analogies:

> Safe eating is like safe sex: You may be eating whatever it was that what you're eating ate before you ate it.
>
> Changing a university has all the difficulties of moving a cemetery.
>
> The juvenile sea squirt wanders through the sea searching for a suitable rock or hunk of coral to cling to and make its home for life. For this task, it has a rudimentary nervous system. When it finds its spot and takes root, it doesn't need its brain anymore, so it eats it! (It's rather like getting tenure.) (Dennett 1991, p. 177)
>
> Bill James on Tim McCarver's book on baseball: "But just to read the book is nearly impossible; it's like canoeing across Lake Molasses."
>
> Red Smith: Telling a non-fan about baseball is like telling an 8-year-old about sex. No matter what you say, the response is "But why?"

In all these cases, there is an analogical mapping that generates surprise and amusement.

In the emotional coherence theory of Thagard (forthcoming), surprise is treated as a kind of metacoherence. When HOTCO shifts from coherent interpretation to another, with units that were previously activated being deactivated and vice versa, the units that underwent an activation shift activate a surprise node. In analogical jokes, the unusual mapping produces surprise because it connects together elements not previously mapped, but does so in a way that is still highly coherent. The combination of activation of the surprise node, the coherence node, and other emotions generates humorous amusement.

Analogies that are particularly deep and elegant can also generate an emotion similar to that produced by beauty. A beautiful analogy is one so accurate, rich, and suggestive that it has the emotional appeal of an excellent scientific theory or mathematical theorem. Holyoak and Thagard (1995, ch. 8), describe important scientific analogies such as the connection with Malthusian population growth that inspired Darwin's theory of natural selection. Thus scientific and other elegant analogies can generate positive emotions such as excitement and joy without being funny.

Not all analogies generate positive emotions, however. Ironies are sometimes based on analogy, and they are sometimes amusing, but they can also produce negative emotions such as despair:

> HONG KONG (January 11, 1998 AF-P) - Staff of Hong Kong's ailing Peregrine Investments Holdings will turn up for work Monday still in the dark over the fate of the firm and their jobs. ...
>
> Other Peregrine staff members at the brokerage were quoted as saying Sunday they were pessimistic over the future of the firm, saddled with an estimated 400 million dollars in debts.
>
> "I'm going to see the Titanic movie...that will be quite ironic, another big thing going down," the South China Morning Post quoted one broker as saying.

Shelley (in progress) argues that irony is a matter of "bicoherence," with two situations being perceived as both coherent and incoherent with each other. The Peregrine Investments-Titanic analogy is partly a matter of transferring the emotion of despair from the Titanic situation to the company, but the irony generates an additional emotion of depressing appropriateness.

The final category of emotion-generating analogies I want to discuss is motivational ones, in which an analogy generates positive emotions involved in inspiration and self-confidence. Lockwood and Kunda (forthcoming) have described how people use role models as analogs to themselves, in order to suggest new possibilities for what they can accomplish. For example, an athletic African American boy might see Michael Jordan as someone who used his athletic ability to achieve great success. By analogically comparing himself to Michael Jordan, the boy can feel good about his chances to accomplish his athletic goals.. Adopting a role model in part involves transferring emotions, e.g. transferring the positive valence of the role model's success to one's own anticipated success, but it also generates new emotions accompanying the drive and inspiration to pursue the course of action that the analogy suggests. The general structure of the analogical inference is:

> My role model accomplished the goal G by doing the action A.
>
> I am like my role model in various respects.
>
> So maybe I can do A to accomplish G.

The inference that I may have the ability to do A can generate great excitement about the prospect of such an accomplishment.

In this paper, I have provided numerous examples of emotional analogies including: analogies about emotions, analogies that transfer emotions in persuasion, empathy, and self-explanation; and analogies that generate emotions in humor, irony, discovery, and motivation. In order to understand the cognitive processes involved in emotional analogies, I have proposed an account of analogical inference as defeasible, holistic, multiple, and emotional. The HOTCO model of emotional coherence provides a computational account of the interaction of cognitive and emotional aspects of analogical inference.

REFERENCES

Barnes, A., & Thagard, P. (1997). Empathy and analogy. *Dialogue: Canadian Philosophical Review, 36*, 705-720.

Blanchette, I., & Dunbar, K. (1997). Constraints underlying analogy use in a real-world context: Politics. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 867). Mahwah, NJ: Erlbaum.

Bosanquet, B. (1920). *Implication and linear inference*. London: Macmillan.

Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithms and examples. *Artificial Intelligence, 41*, 1-63.

Fenno, R. F. (1978). *Home style: House members in their districts*. Boston: Little, Brown.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT Press/Bradford Books.

Holyoak, K. J., Novick, L. R., & Melz, E. R. (1994). Component processes in analogical transfer: Mapping, pattern completion, and adaptation. In K. J. Holyoak & J. A. Barnden (Eds.), *Advances in connectionist and neural computation theory, Vol. 2: Analogical connections*. (pp. 113-180). Norwood, NJ: Ablex.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press/ Bradford Books.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A-parallel constraint-satisfaction theory. *Psychological Review, 103*, 284-308.

Lockwood, P., & Kunda, Z. (in press). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology.*

Meichenbaum, D. (1994). *A clinical handbook/ practical therapist manual for assessing and treating adults with post-traumatic stress disorder (PTSD).* Waterloo, Ontario: Institute Press.

Mill, J. S. (1970). *A system of logic.* (8 ed.). London: Longman.

Salmon, W. (1984). *Logic.* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Shelley, C. P. (forthcoming). Multiple analogies in evolutionary biology.

Shelley, C. P. (in progress). Irony.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435-467.

Thagard, P. (1992). *Conceptual revolutions.* Princeton: Princeton University Press.

Thagard, P. (forthcoming). Emotional coherence: Trust, empathy, nationalism, weakness of will, beauty, humor, and cognitive therapy. *unpublished.*

Thagard, P., Eliasmith, C., Rusnock, P., & Shelley, C. P. (forthcoming). Knowledge and coherence. In R. Elio (Ed.), *Common sense, reasoning, and rationality* (Vol. 11,). New York: Oxford University Press.

Thagard, P., Gochfeld, D., & Hardy, S. (1992). Visual analogical mapping, *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 522-527). Hillsdale, NJ: Erlbaum.

Thagard, P., & Millgram, E. (1995). Inference to the best plan: A coherence theory of decision. In A. Ram & D. B. Leake (Eds.), *Goal-driven learning:* (pp. 439-454). Cambridge, MA: MIT Press.

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science, 22,* 1-24.

ANALOGY AS A COGNITIVE VEHICLE IN PENETRATING A NEW DOMAIN

Adam Biela
Catholic University of Lublin

1.INTRODUCTION

According to Heraclit one can not enter twice the same river, because the river is **a new one**, different than it was before. This is why the ancient philosophers proposed the conception of **panta rei**. It is not only the psychologists who could say that a need for seeking a novelty or a need for a change is one of the central human desires. The external control and human being himself or herself is changing its, his or her state into the new one. We are still coping with the changing environment. Let us consider the common verbal expressions, which deal with the verb "new". The dictionary expressions refer the meaning of "new" as follows: - recent in origin; novel; not known before; different; unaccustomed; fresh after any event; not second hand (see: The University English Dictionary edited by R. F. Patterson). - not existing before; lately discovered or invented; recently born (see: The Family Dictionary edited by Collins).

- different / **a whole new „ball game"**; a separate issue or matter very different from the matter under discussion; a new situation very different from the present one; (see: English idioms edited by Oxford University Press p.66)

- **be new to the game** - lack of experience in an activity, job or situation (p.155)

- **new blood** - someone new to an organisation, job or work who is expected to bring new ideas, innovations (p.214)

A human desire to search for the new word and to cope with novelty can be seen in such verbal expressions, which promoted the new streams of history, discoveries and civilisations. The examples might be:

New Style - a chronological term to demote dates reckoned by the Gregorian calendar;

New Deal - a campaign initiated in 1933 by President Franklin Roosevelt in USA involving a complete overhaul of American economic life, the development of the national resources and the safeguarding of conditions for labour; New Learning - the Renaissance;

New Testament - later of the two main divisions of the Bible;

New World - North and South America;

The above expressions denote the notion "new" as an unknown reality, different than the well-known and experienced before. The new situation is a reality which is in question because it is at least less known. The question is how to cope with the new reality which, on one hand, expected to be reached, and, on the other hand, is risky because of its novelty and requires decision which way to go and how to "possess" cognitively the current stream of the environment.

2. COGNITIVE COPING WITH NEW SITUATION

Cogito ergo sum in a new situation requires to cope cognitively with unknown and uncertain environment. Cognitive coping with new environment assumes to employ schema of inference and forecasting not only to survive but also to develop human potentiality. Generally speaking there are also two schema's which could be used to cope with new situations: (1) deductive reasoning schema's based on logical implication connection; and (2) analogical reasoning which is not based on logical implicational foundation. Unfortunately, deductive inference can not be employed in many new situations where general premises are hardly to be formulated. In those cases analogical reasoning can be only employed to draw conclusions about the situations which are in question. This is why analogical inference is used to be called as reasoning from case to the case. In concluding our reflection on cognitive coping with new situation we can say that analogy can be recognised as cogito which leads coping with the unknown environment.

Analogical cogito is a cognitive "vehicle" for searching connections, relations, correspondence between the new domain which is in question and the well known domain. However, to be more precise we should state that analogy assumes a comparison between two domains, situations, fields or areas in respect with the specific relations. Therefore analogy can be interpreted as a cognitive schema i.e. a kind of mental principle or human mind's structural path and at the same time a mental vehicle which drives for searching relational connections (or correspondence) within the considered domains and between them.In a more formal way we can define analogy (Anal) as a two compound complex relation expressed as:

(1) Anal = R(D, D') or

(2) Anal = DRD', where

D - a well known domain, situation;

D'- a new (less known) domain, situation, i.e. which is in question. The relation considered in (1) or (2) is a complex one because its compounds domains D and D' correspond one to another with respect to the constituting their relations, R_n and R'_n respectively:

(3) $D = R_n(x_1,x_2,...,x_i,...,x_n)$

(4) $D' = R'_n(x'_1,x'_2,...,x'_i,...,x'_n)$, where

R_n - a base relation for analogy which denotes that the known domain D corresponds to the new domain D' in such a way that the relation R_n constituting the D fits the relation R'_n constituting the D';

R'_n - a relation constituting the new domain D' (which is in question) and corresponding to the base relation for analogy R_n constituting the known domain D;

$x_1,x_2,...,x_i,...,x_n$ - the compounds of the base relation for analogy R_n

$x'_1,x'_2,...,x'_i,...,x'_n$ - the compounds of the relation R'_n corresponding to the base relation for analogy R_n.

After completing (3) and (4) we can formally define analogy in a more complex formula, respectively:

(5) $\text{Anal} = R[R_n(x_1,x_2,...,x_i,...,x_n), R'_n(x'_1,x'_2,...,x'_i,...,x'_n)]$,

(6) $\text{Anal} = [R_n(x_1,x_2,...,x_i,...,x_n)] R[R'_n(x'_1,x'_2,...,x'_i,...,x'_n)]$.

Analogy is used as a scheme to cope with a problem in a new domain, situation. This scheme allows to formulate two premises and to draw conclusion concerning the unknown compound of the base relation within the new situation:

Premises:

1. The domain D' corresponds to the domain D with respect to the constituting them relations R'_n and R_n, accordingly.

2. The compound x'_i of the relation R'_n is unknown in the domain D' but the others are known and fit the corresponding compound, of the relation R_n in the domain D.

(7) Conclusion: The x'_i is like x_i .

The first premise of the scheme (7) states a base for analogy, i.e. a correspondence between the compared domains with respect to the appropriate relations. The second premise says that one compound of the relation constituting the domain which is in question, is unknown while the others fit the corresponding compounds of the relation constituting the known domain. Therefore, the analogical conclusion completes the correspondence between the relations R'_n and R_n stating that the unknown compound x'_i has found its corresponding compound x_i.

3. COGNITIVE VEHICLE

Analogy is used as a cognitive vehicle in science, particularly when:

- formulating new hypothesis;
- introducing new concepts;
- arguing new statements.

Analogy plays also a role of a cognitive path in economy, politics or social endeavour when the participants of economic, political and social life are facing problems in new situations and particularly in transformation of a macrosystem after the collapse of the totalitarian system called the communism.

Let us consider now more general statements concerning employing analogy as a cognitive schema in the proposition of solving problems in a transformation situation in Central and Eastern European countries.

Generally, one can say that D' is a new unknown situation, when he or she is facing problems dealing with:

- restructuring of centrally managed and state-owned economy into a private sector which better fits the realities of free market;
- legislation which deals with Parliament activity and then with executing the law by the government;
- building democratic infrastructure which enables citizens to participate in social economic and political life.

What are the known domains D which could help to find an analogy vehicle to search for corresponding schema's, structures, methods, law regulations, institutions that are appropriate to cope with the actual problems. As the potential domains to look for economic, legislative or political analogies in the Polish transformation situation, could be considered as situations which deal with market economy experiences, democratic institutions and legislation are:

(a) the period of the pre-IInd World War Poland, i.e. since 1918 when Poland became an independent state after the Ist World War which finished the partition of Poland until September the 1st 1939, i.e. the beginning of the German occupation and then the Soviet occupation;

(b) the West European, American or Asiatic market economy institutions and solutions;

(c) the democratic and free market economy solutions known in some local communities, regions or countries which could be treated as leading or good examples of macrosystem transformation in post-communist countries.

Therefore the considered analogies are called the pre-war Polish analogies or the contemporary West European, American or Asiatic analogies. As far as the content, object or the goal of the analogical schema is specified we are facing the defined analogies: legislative, institutional, behavioural, infrastructural, organisational - respectively. If the known domain (situation) is actually existing, the appropriate leading analogy can be called actual or contemporary (local, regional, domestic or foreign analogy). If the known situation which leads the analogy can be learned only from the past, the appropriate analogy can be named as a historical one.

Analogy as a cognitive vehicle towards new economic behaviour, new market economy institution, new legislative solution can have a strong background or can be supported by some surface or superficial base. This means that the base relations for analogy, i.e. R_n and R'_n could be substantial or accidental.

Biela (1993) formulated six conditions of analogical correspondence of the relations which fulfilment seems to be relevant to the validating inference based on analogical connection. The condition related to the substantives of the base relation is expressed as **the constitutiveness condition**. It states that, if domains D and D' are sufficiently precisely defined and the relation: R_n and R'_n are as well, then, according to the available level of scientific knowledge, the existence of the domain D without the relation R_n and the existence of the domain D' without the relation R'_n is impossible. The meaning of „existence" depends here on the type of domain in question and the accepted concept of the domain being considered. For example, the mode of existence of the mathematical domain depends on the assumed philosophy of mathematics. The same is true of the fine art or music domains where their mode of existence depends on the assumed theory of the fine art work or music composition. The existence of the natural science domains depends also on the accepted philosophy underlying modern theory of the particular discipline, e.g. within the theory of physics could be considered the discussion between the Duhemist and Campbellian approach (see: Hesse, 1963). In the social sciences „existence" depends mainly on social perception of the relation which is in question. If we create the sentential functions f(D) and f(D') from the respective nominal symbols D and D', this condition might be expressed in the following way (the condition of constitutiveness):

(8) $[R_n \Rightarrow f(D)] \cap [\sim R_n \Rightarrow \sim f(D)]$ and

(9) $[R'_n \Rightarrow f(D')] \cap [\sim R'_n \Rightarrow \sim f(D')]$.

The condition emphasised that not all relations recognised within the domain D and D' could be stated as the base for analogy, as many of them are not constitutive for these domains (i.e. it is still possible to see the respective domain without involving many various relations). If an analogy connection was to be based on surface relations that do not fulfil the condition (i.e. ones that are not constitutive for the domain D and D') then the inference based on such a connection will not guarantee any valuable result. And, moreover, such kinds of surface connections are hazardous because they create only the appearance of rational thinking by analogy.

If the condition of constitutiveness is considered in the applied areas of social sciences we should analyse as a criterion the social perception of distributive justice in the specific field of endeavour. To be more specific, the social perception of risk and benefits analysis should be considered with respect to the issue which is in question. Therefore, as far as economic transformation is considered, the specific questions are:

1. What are the risk and costs of the transformation?

2. Who is the beneficier of the transformation in post-communist countries?

3. Who is taking risk and paying the main cost of this process?

4.EXAMPLES OF ANALOGIES

Let us consider some examples of analogies employed in macrosystem transformation time in Poland.

First group of transformational analogies are the coping mechanisms which employ some surface behavioural or institutional analogies.

Conserving old structures under a new coat of paint

A frequent coping strategy is to hold and to conserve old organisations, institutions and structures while attempting to adopt them to new political and constitutional circumstances. The adaptational level here is very superficial. This kind of adaptation could be metaphorically described as „painting over a heavily rusted car". It is a case where the old political party, central economic institutions, and local municipal governments want to survive under the new political and economic circumstances. Therefore, they decide to make some cosmetic changes such as a new name, a surface reorganisation, minimal reduction of employees, changing leaders, etc., without any serious intention to change the deep structure of the institution or reformulate its goal and function in the new environment.

The behavioural, institutional and organisational analogy is based on the conserved behavioural patterns and institutions learned and structured during the centrally managed economy period. This is a conservative analogy which really avoids serious transformation.

Constructing new institutions according to old patterns

Another form of „surface" mental adaptation is constructing a new alternative, and formally independent institution according to old patterns. This sounds paradoxical, but often these old patterns were criticised just by the people who form new institutions. These patterns deal mainly with monopolistic and totalitarian - centralist mind - sets. This happens quite often in newly installed political parties, new local administrative centres, central governmental institutions, etc. The point here is that people are not able to behave in a new way, even if they create a new institution. Analogy here also is based on behavioural and organisational patterns learned in the climate of totalitarian mentality. This is also a conservative analogy which secures a continuation of the mental climate in the new political circumstances. Some critical time is needed to change the old patterns in people's behaviour. It requires to change the base relation in analogical reasoning.

Ersatz standards of freedom and high living

Another coping mechanism is to find some available evidence of freedom or a high standard of living that could play the role of an ersatz, substituting for real freedom or a real improvement in the standard of living. Examples of substitutes playing such a role could include unusual goods like Western-style clothes (even if second-hand), used Western cars (even if rusted), or Western style sex-shops and sex-magazines. These substitutes create an illusory atmosphere that the desired changes have already succeeded. The cognitive mechanism leading such behaviours can be called as an ersatz analogy which is based on very superficial, easily available and of immediate gratification behavioural effect. This kind of analogy allows easily to achieve a sense of an illusory participation in a Western high standard of living. Facing political changes, some people prefer immediate gratification instead of waiting for the long-term, delayed effects of the changes. Such people prefer having ersatzes, which can be achieved in a short time instead of the real desired changes themselves. Ersatz analogies touch surface and superficial relations of the Western life which can not be stated as a reality of the market economy world. Unfortunately they create an illusory atmosphere that the desired transformation related with Western democracy and free market economy have already succeeded.

5. LEGISLATIVE ANALOGY

A good example of using historical analogy which employs the experiences of the pre-war Poland, is an initiative of restitution of i.e. Prokuratoria Generalna which is the institution controlling the managing of the State Treasure. Let us state the background of the Prokuratoria Generalna legislation analogy.

(a) The known situation (S).

The known situation (S) is here the pre-war period when the institution of the Prokuratoria Generalna was introduced by the legislation of the Polish Sejm in 1919. The legislation was initiated at the very beginning of the II Polish Republic by the Parliament. The architects of the Polish state believed that building of market economy required an institution of a very high professional and moral authority to control the efficiency of managing of the national treasure resources. The pre-war Polish economy reached significant development in terms of its potentiality, level of investment, stock market infrastructure, macroeconomic indicators. The pre-war Polish Prokuratoria Generalna functioned efficiently and reached a high professional prestige and moral authority.

(b) The new situation (S').

The designers of the Polish macroeconomic transformation after a collapse of communism are facing difficulties in building market economy. However, the bigger problem is more how to restructure the state-owned enterprises into the private companies which can cope with market reality and in the same time fit the Polish economy long-term benefits perspective. Building the market economy in Poland based on privatisation requires the institution to control the process at the very beginning of building market economy in the III Polish Republic. Therefore the Polish Sejm at the very beginning of its IIIrd cadence articulates a legislative initiative for the Prokuratoria Generalna which resembles the pre-war institution of the same name. Moreover, the legislative proposal of March 1998, in the Article 1 refers to the pre-war tradition of this kind of institution.

6. PRIVATISATION ANALOGIES

The process of transformation in post-communist countries drives towards changing the ownership status of the state-owned enterprises into the private entities. However the problem is who should be the owner of the enterprises and which model of privatisation to choose. The Polish way of privatisation employs a variety of models which are based on analogies drawn from comparing the West European or the USA private companies with the Polish companies of the same branch. The proposed models deal with such forms of privatisation as:

- capital privatisation;
- Employee Stock Ownership Plan (ESOP);
- leasing form.

The mentioned above forms of private companies are the well known domains which inspire to think by analogy about possible ownership transformation of the related Polish companies. However, it is obvious that the macroeconomic, social and political environment of the Western companies is hardly similar to the Polish ones. Moreover the risk of the Polish transformation is in it that the changes are sudden and in a large scale what was never in the Western world the case where the development of private companies took years. Nevertheless, analogies are the mental bridges which lead to solving the Polish problems of privatisation.

The Western models of the private ownership can not be applied literally into the Polish situation. They can be employed partially. Let us consider the example on the American ESOP. For the same Polish companies the ESOP analogy became a direct model for privatisation. However, the Polish privatisation requires more extensive model for a large-scale privatisation where the participants will be the Polish citizens whose insufficiently paid work was accumulated into the investment capital. This is why they have a right to participate in a privatisation of the state-owned companies. This kind of privatisation is called in Poland **the Program Powszechnego Uw3aszczenia** what might be translated as **the Citizens Ownership Program** (see: Sejm print No 400). This program gives a chance to the Polish citizens to participate in the ownership and play an active role in the allocation of the investment capital. The program uses the instruments and institutions of the stock market. The intention of this program is, among others, to concentrate the local and the regional capital within the Local Mutual Investment Funds. The idea of such capital institutions were drawn by analogy to (a) the Western Mutual Investment Funds, and to (b) the pre-war Polish local Saving Cooperatives (called the Kasy Stefczyka - from the name of their promoter).

7. REMARKS

Analogy is the most intriguing cognitive principle which allows to draw conclusion, particularly in new areas, domains and situations. However, the value of the drawn analogical conclusion depends on the relation which is called a base for analogy. In other words, analogical reasoning might be founded on surface, superficial base relations or on substantial background.

Building new economy and democratic institutions after a collapse of totalitarian system requires not only a mental adaptation into a new situation but shaping and restructuring the situation which is in question. Analogical reasoning plays an important role both in mental adaptation and in shaping new situation. However, the participants of transformation use more or less sophisticated analogies in coping with new political, social and economic environment.

8. REFERENCES

Biela, A. (1991). Analogy in science.Frankfurt am Main: Peter Lang.

Biela, A. (1993). Psychology of analogical reasoning.Stuttgart: S.Hirzel Verlag Stuttgart.

Hesse, M. (1963). Models and analogies in science.London: Sheed and Ward.

EXPLORING ANALOGY IN THE LARGE

Kenneth D. Forbus
Institute for the Learning Sciences, Northwestern University
1890 Maple Avenue
Evanston, IL, 60201
Email: forbus@ils.nwu.edu

Significant progress has been made in creating cognitive simulations that model a variety of phenomena in analogy, similarity, and retrieval [1]. To date, most models have focused on exploring the fundamental phenomena involved in matching, inference, and retrieval. While there is still much to be discovered about these areas, the time seems right for more energy to be focused on simulating the roles analogy places in larger-scale cognitive processes: what I call *large-scale analogical processing*.

Psychological evidence suggests that structural alignment plays a central role in many cognitive processes [c.f. 2,3,4,5]. An important challenge for cognitive simulations is that they be capable of modeling the same breadth of phenomena. Exploring these issues requires moving beyond simulating isolated modules and working in toy domains to creating larger-scale simulations that model a wider range of cognitive phenomena. In addition to cognitive modeling, we believe that the state of the art in analogical processing has advanced to the point where it can be used to create fundamentally new kinds of applications.

This talk describes two examples of how we are using cognitive simulation to explore the roles of structure-mapping [6] in large-scale analogical processing. We use the Structure-Mapping Engine (SME) [7,8,9] to model the comparison process that underlies analogy and similarity, and MAC/FAC [10,11] to model similarity-based retrieval. The examples are:

- A design coach for students learning engineering thermodynamics that is accessible via email [12]. Students use CyclePad, an articulate virtual laboratory [13] for engineering thermodynamics, to create designs for power plants, refrigerators, and other systems. A built-in email facility enables them to ask for help from an automatic server, including advice on improving their designs. The design coach uses MAC/FAC to retrieve cases and uses SME to create advice showing how the transformation in the case can be tailored to a student's design. By using SME and MAC/FAC and our tools, human domain experts can add cases to the library without hand-coding representations or handindexing them for retrieval.
- An account of mental models we are developing to help explain common sense reasoning about the physical world [14]. Two common explanations for qualitative mental models are high-resolution imagery and first-principles reasoning from general domain theories. We propose instead *similarity-based qualitative simulation* as a psychologically plausible mechanism for common sense prediction tasks. Similarity-based qualitative simulation uses analogical retrieval and mapping of qualitative representations to make predictions in novel situations based on previously experienced behaviors.

REFERENCES

1 Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy: Introduction. *American Psychologist, 52*, 32-34.

2 Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45-56. (To

be reprinted in *Mind readings: Introductory selections on cognitive science,* by P. Thagard, Ed., MIT Press)

3 Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrieval from inferential soundness. *Cognitive Psychology, 25,* 524-575.

4 Markman, A. B., & Gentner, D. (1993c). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431-467.

5 Gentner, D., Brem, S., Ferguson, R. W., Wolff, P., Markman, A. B., & Forbus, K. D. (1997). Analogy and creativity in the works of Johannes Kepler. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 403-459). Washington, DC: American Psychological Association.

6 Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7, 155-170.

7 Falkenhainer, B., Forbus, K., and Gentner, D. (1986, August) The Structure-Mapping Engine. *Proceedings of AAAI-86*, Philadelphia, PA

8 Falkenhainer, B., Forbus, K., Gentner, D. (1989) The Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence*, 41, pp 1-63.

9 Forbus, K., Ferguson, R. and Gentner, D. (1994) Incremental structure-mapping. *Proceedings of the Cognitive Science Society,* August.

10 Gentner, D., & Forbus, K. D. (1991, August). MAC/FAC: A model of similarity-based access and mapping. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society,* Chicago, IL.

11 Forbus, K., Gentner, D. and Law, K. (1995) MAC/FAC: A model of Similarity-based Retrieval. *Cognitive Science,* 19(2), April-June, pp 141-205.

12 Forbus, K., Everett, J.O., Ureel, L., Brokowski, M., Baher, J., and Kuehne, S. (1998) Distributed Coaching for an Intelligent Learning Environment. *Proceedings of the Twelfth International Workshop on Qualitative Reasoning.* Cape Cod, Mass. (Proceedings available as a AAAI Technical Report)

13 Forbus, K. Using qualitative physics to create articulate educational software. *IEEE Expert*, **12**(3), May/June 1997.

14 Forbus, K., & Gentner, D. (1997). Qualitative mental models: Simulations or memories? *Proceedings of the Eleventh International Workshop on Qualitative Reasoning*, Cortona, Italy.

COMPARISON AND COGNITION

Dedre Gentner
Northwestern University

Similarity, metaphor and analogy are fundamental mechanisms of learning. In this research I suggest a unified framework of structural alignment between situations or domains that highlights common structure and allows further properties to be projected. This structure-mapping framework suggests notions of structural consistency, systematicity and candidate inferences that offer new insights into how comparison is used to perceive commonalities and differences, project inferences and derive new abstractions.

One advantage of this framework is that it allows us to model extended metaphors that map large-scale belief systems. In one series of studies, we tested whether extended metaphors are processed as mappings from one conceptual system to another (Gentner & Boronat, 1992; in preparation). We gave participants a series of consistent metaphoric statements from one domain (the base) to another (the target); they read these statements one at a time on a computer screen. Half the subjects were in the consistent condition, and received a metaphor that remained consistent throughout the passage. The other half received a different metaphor, so that the mapping shifted at the last sentence.

e.g.,

CONSISTENT MAPPING [mind as knife - mind as knife]

"...After just three hours she had lost her edge...

Her mind was too dulled with fatigue for her to think well."

INCONSISTENT MAPPING [mind as engine — mind as knife]

"...After just three hours she had run out of steam...

Her mind was too dulled with fatigue for her to think well."

As predicted by the domain-mapping hypothesis, participants were slower to read the last sentence when there was a shift in the underlying mapping.

However, this was only true for novel metaphoric phrases. The processing of conventional metaphoric phrases was not disturbed by the shift in mapping. This finding would be predicted by Bowdle and Gentnerfs (in preparation) career of metaphor hypothesis, that metaphors are initially processed as mappings, but eventually become processed as lexical word senses (See also Gentner & Wolff, 1997). The implication of this finding is that structure-mapping processes are used to understand novel metaphors, and further that these processes can serve to create new word meanings.

I suggest that alignment and mapping processes are a major force in human learning and development. Analogical mapping promotes learning and conceptual change in three ways: by inviting inferences from one situation to the other, by promoting schema abstraction across the two situations, and by prompting re-representation of one or both situations. I will present evidence from studies of children and adults to show that comparison processes are a major mechanism of spontaneous learning and a natural route towards abstract systems of understanding.

In summary, my thesis is that analogical thinking is fundamental to human cognition.

BIBLIOGRAPHY

Bowdle, B. & Gentner, D. (1995, November). The career of metaphor. Paper presented at the meeting of the Psychonomics Society, Los Angeles, CA.; mss. in preparation.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine:

Algorithm and examples. Artificial Intelligence, 41, 1-63.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7, 155-170.

Gentner, D., & Boronat, C. B. (1992). Metaphor as mapping. Paper presented at the Workshop on Metaphor, Tel Aviv.

Gentner, D., & Boronat, C. B. (in preparation). Metaphors are (sometimes) processed as generative domain-mappings.

Gentner, D., & Markman, A. B. (1997). Structure-mapping in analogy and similarity. American Psychologist, 52, 45-56. (reprinted in Mind readings: Introductory selections on cognitive science, by P. Thagard, Ed., MIT Press)

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. Cognition, 65, 263-297.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), Perspectives on language and thought: Interrelations in development, (pp. 225-277). London: Cambridge University Press.

Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. Journal of Memory and Language, 37, 331-355.

ANALOGY IS LIKE COGNITION: DYNAMIC, EMERGENT, AND CONTEXT-SENSITIVE

Boicho Kokinov

Department of Cognitive Science
New Bulgarian University
21, Montevideo Str.
Sofia 1635, Bulgaria

Institute of Mathematics and Informatics
Bulgarian Academy of Science
Bl. 8 Acad. G. Bonchev Str.
Sofia 1113, Bulgaria
kokinov@cogs.nbu.acad.bg

ABSTRACT

This paper presents several challenges to the models of analogy-making, namely the need for building integrated models, the need for using dynamic and emergent representations, the need for using dynamic and emergent computation, and the need to integrate analogy-making with other cognitive processes. Some experimental data are reviewed which substantiate these needs and the main ideas how the AMBR model of analogy-making could meet these challenges are presented.

1. FROM THE ANATOMY TOWARDS THE PHYSIOLOGY OF ANALOGY-MAKING: THE NEED FOR INTEGRATED AND DYNAMIC MODELS

For a long time now the research on analogy has concentrated on the anatomy of analogy-making, i.e. on decomposing it into pieces (representation building, retrieval, mapping, transfer, evaluation, learning) and trying to understand how each individual piece works. A number of successful models of various subprocesses (mainly of mapping and retrieval) have been built which account for most of the psychological data and make useful predictions: SME and MAC/FAC (Gentner, 1983, Falkenheiner, Forbus, Gentner, 1986, Forbus, Gentner, Law, 1995), ACME and ARCS (Holyoak, Thagard, 1989, Thagard, Holyoak, Nelson, Gochfeld, 1990, Holyoak, Thagard, 1995), IAM (Keane, Ledgeway, Duff, 1994), etc.

The big challenge in modeling analogy-making (and human cognition in general) is to move on from the atomistic and analytical approach of Democritus (469-370 BC) towards the holistic and interactionist approach of Heraclitus (544-481 BC), i.e. to start building integrated models of the phenomenon as a whole. These models should unite contraries and account for data arising from the interaction between subprocesses, which cannot be explained by an isolated model of a subprocess. Such models are gradually emerging. Thus the CopyCat and TableTop models (Hofstadter, 1995, Mitchell, 1993, French, 1995) integrate representation building with mapping and transfer, LISA (Hummel and Holyoak, 1997) integrates access, mapping, transfer, and learning. AMBR (Kokinov, 1988, 1994c) integrates access, mapping, and transfer.

Heraclitus took the view that "Everything flows, everything changes", i.e. the dynamics of change is more important and informative than static objects and states. This is the next challenge to the current models: they should explain and predict not only the outcomes of the analogy-making process but also its dynamics. Unfortunately, only scare data is available on the dynamics of the process. This means that such data will have to be gathered by using experimental paradigms extensively used in other domains, for example, on-line experiments measuring reaction times, analysing thinking-aloud protocols, etc. These methods have already been used in analogy research but to a very limited extent (Ross and Sofka, 1986, Keane, Ledgeway, and Duff, 1994, Schunn and Dunbar, 1996).

There are already experimental data which support the existence of interaction effects between the subprocesses of analogy-making. Thus Keane, Ledgeway, and Duff (1994) have demonstrated a very strong ordering effect, i.e. effect of the order of presentation of the target problem elements on the response time for solving the problem. Thus in the "singleton-first" condition subjects found the mapping twice as fast as subjects in the "singleton-last" condition. These data can be considered as evidence for the interaction between perceptual and mapping processes. It would be even more interesting to find the reverse patterns: the mapping already established facilitating the perception of certain elements.

The analysis of thinking-aloud protocols done by Ross and Sofka (1986) revealed that the retrieval of various elements of the source domain is interrelated with the mapping between the two domains, i.e. the already established mappings guide the retrieval of specific source elements. These data cannot be explained by a serial model of analogy-making where first the source is being retrieved and then the source and target are mapped. An extensive discussion of this phenomenon and its modeling in AMBR as well as simulation data obtained with AMBR can be found in (Petrov, Kokinov, this volume). AMBR predicts also the reverse influence: the specific order of retrieval of elements of the source domain will facilitate certain mappings. As a result of these interactions, a pattern of retrieval has been demonstrated where initially one source domain looks more promising and is better retrieved based on the greater superficial similarity, but as soon as mapping starts (in parallel to the continuing retrieval of domain elements), the higher structural correspondence between a second source domain and the target and the established mappings make it possible for the second domain to be ultimately better retrieved and mapped which would be impossible if the retrieval and mapping were sequential isolated and irreversible processes.

Finally, a study currently underway involves video recording of subjects solving a formatting task on a computer screen. The video protocols demonstrate a complex interaction between perceiving elements on the screen (including figure/background perception), retrieving elements from memory, mapping between these elements, and performing actions on the screen, the results of which are further perceived and mapped to expectations.

The explanation of all these data requires models which abandon the serial type of processing and which move on towards parallel processing which will allow the various subprocesses to interact dynamically with each other. AMBR is one such model that is based on the highly parallel cognitive architecture DUAL (Kokinov, 1994a, 1994b). All processes in AMBR are running in parallel and interacting with each other. Moreover, as described in section 3, each of these subprocesses emerges from the collective behavior of many micro-agents and thus is also inherently parallel. Since the micro-agents are taking part in various subprocesses there are no clear-cut boundaries between the various processes themselves.

Before the dynamics of computation in AMBR can be presented, the need for dynamic representations that will change in the course of analogy-making will be discussed in the next section.

2. FROM PRINTED TEXT TOWARDS MOVING PICTURE: THE NEED FOR DYNAMIC AND EMERGENT REPRESENTATIONS

A printed text is a static representational object while a moving picture is a dynamic representation which emerges from the continuously changing frames. Moreover, this dynamic representation does not exist physically (only the static frames exist physically), it exists only in our consciousness. Analogously, memory traces may be considered either as physically existing static entities, or as emergent phenomena which are constructed in our consciousness.

From the very beginning of memory research the view of memory as consisting of stable representations has been under fire. Thus Bartlett (1932) has shown that episodes are grouped into schemas and their representations are systematically shifted or changed in order to fit these schemas. Research on autobiographical memory has provided evidence that people modify their memories by dropping elements (schematising), including new elements (filling in), replacing elements (distorting), etc. Loftus (1977, 1979) has convincingly demonstrated a number of interference effects. One example involves subjects looking at a movie where a blue car does not stop at the site of an accident. Later on in a questionnaire a number of questions are asked about a different green car. As a result, when asked about the color of the car which did not stop, subjects are quite confident that it was green. In another study subjects claim they have seen broken glass in a car crash whereas there was no broken glass in the movie shown to them.

Neisser and Harsch (1992) have demonstrated that the so-called "flash-bulb memory" does not exist but that descriptions constructed by human memory are so vivid that people strongly believe they are true. One day after the *Challenger* accident they asked subjects to tell them (and write down) how they learnt about the accident: whether they heard it on the radio, or saw it on TV, or learnt it on the street, in the supermarket, from friends. They asked further the subjects in the study what they were doing when they learnt about the accident, what their reactions were, etc. One year later the experimenters asked the same subjects whether they still remember the accident and how they learnt about it. People claimed they had very vivid ("flash-bulb") memories about every single detail and they started to tell the experimenters a very different story from the one they told before. Even after the experimenters showed them their own writings they could not believe that the new story they were telling the experimenters was not true.

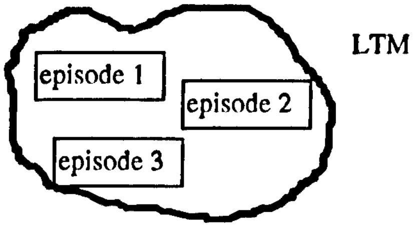


Figure 1. Centralized and frozen representations of episodes in LTM.

Although it has long been demonstrated that human memory is a (re)constructive device rather than a store of stable memory traces from our past, models of analogy-making tend to ignore that fact. Typically these models would have a collection of representations of past episodes (prepared by the author of the model) "stored" in long-term memory (LTM), one or more of which would be "retrieved" during the problem solving process and would serve as a base (or source) for analogy. The very idea of having singular centralized and frozen representations of base episodes is at least questionable, but it underlies most analogy-making models, and certainly all case-based reasoning systems (Figure 1).

Research on retrieval in analogy-making has concentrated on how people select the most appropriate episode from the vast set of episodes in LTM. It has been established that the existence of similar objects, properties or relations in the two domains is the crucial factor for retrieval (Holyoak & Koh, 1987, Ross, 1989) and that is why remote analogies are very rare. On the other hand, structural similarities can also facilitate retrieval under certain circumstances, when there is a general similarity between the domains or story lines (Ross, 1989, Wharton, Holyoak, Lange, 1996). There is not much research either on the dynamics of the process of retrieving (or constructing), or on how complete the resulting descriptions of the episodes are.

A recently conducted experiment was designed as a replication of Holyoak and Koh's (1987) Experiment 1. However, a thinking-aloud method was used. Subjects discussed the solution of the radiation problem in a class on thinking within an introductory Cognitive Science course. From 3 to 7 days later they were invited by different experimenters to participate in a problem-solving session in an experimental lab. They had to solve the light bulb problem. Almost all subjects (except one who turned out not to have attended the class discussing the tumor problem) constructed the convergence solution and explicitly (in most cases) or implicitly made analogies with the radiation problem. We were interested how complete and correct their spontaneous descriptions of the tumor problem story were. It turned out that remembering the radiation problem is not an all-or-nothing case. Different statements from the story were recollected and used with varying frequency. Thus the application of several X-rays on the tumor was explicitly mentioned by 75% of the 16 subjects participating in the experiment, the statement that high intensity rays will destroy the healthy tissue was mentioned by 66% of the subjects, while the statement that low intensity rays will not destroy the tumor was mentioned only by 25%. Finally, no one mentioned that the patient would die if the tumor was not destroyed. All this demonstrates a partial retrieval of the base: which elements of the base will be retrieved depends on the pragmatically important aspects of the target problem.

On the other hand, there were some insertions, i.e. "recollections" of statements that were never made explicit in the source domain description. Thus one subject said that the doctor was an oncologist which was never explicated in the radiation problem description (nor should it be necessarily true). Another subject claimed that the tumor had to be burnt off by the rays, which was also never formulated in that way in the problem description.

Finally, there were borrowings from other possible bases in memory: thus one subject said that the tumor had to be "operated by laser beams" while in the base story the operation was even forbidden. Such blendings were very frequent between the base and the target, thus 7 out of the 11 subjects spontaneously re-telling the base (the radiation) story were mistakenly using laser beams instead of X-rays to destroy the tumor. This blending is evidently the result of the correspondence established between the two elements and their high similarity.

In summary, the experiment has shown that remindings about the base story are not all-or-nothing events and that subjects make omissions, insertions, and blendings with other episodes.

The representation of episodes in AMBR is de-centralized, which means that separate elements of the episode's description are represented by separate memory elements (called micro-agents in the DUAL cognitive architecture). Thus the episode as a whole is represented by a coalition of agents, but there is no guarantee that the whole coalition will be activated and become part of WM. Depending on the weights of the links between the agents the coalition might be looser or tighter. This makes it possible to model the above mentioned psychological effects. Thus very often only part of the agents in a coalition are being activated above the Working Memory (WM) threshold and thus the corresponding episode is only partially retrieved. Depending on the retrieval cues used various partial recollections will be produced.

Blendings also happen in AMBR. Thus agents representing aspects of several different episodes can be concurrently activated in WM. Mappings between elements of the target and elements of all partially retrieved episodes can be established in parallel and compete with each oth-

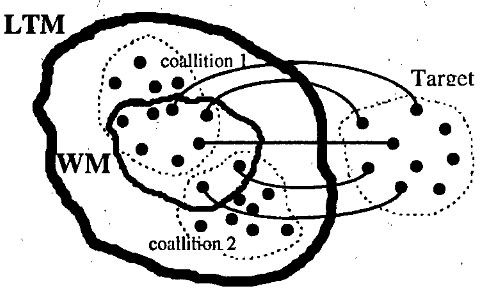


Figure 2. Blending of two episodes (represented by two coalitions) which are partially retrieved in WM and partially mapped on the target coalition. (The target coalition is also part of WM, but is depicted separately for simplicity of the diagram).

er. Typically the support that the agents in one coalition receive from each other is enough to achieve a global emergent "winner" episode. However, in some cases one or more aspects needed for the mapping (having counterparts in the target) are missing in the representation of an episode, or are not retrieved in WM, but instead corresponding elements from other episodes are retrieved. In such a case a blending between the episodes can happen, i.e. the target elements are partially mapped to elements of one base and partially to elements of another base (Figure 2).

Finally, insertions (analogous to the doctor-oncologist case) are also possible in AMBR. Semantic knowledge is represented in a similar decentralized fashion, i.e. different aspects of a concept are represented by different agents. Suppose, for example, that there is a general rule saying that liquids are typically held in containers. Suppose now that an episode is being retrieved in which water is heated by an immersion heater. It might well be the case that the fact that the water was in a glass was either not encoded at all, or was not retrieved under the current circumstances. At the same time the target situation involves tea being heated in a pot on a plate. The agent representing the fact that the tea is in the pot will activate many agents representing similar facts and in particular the one representing liquids being in containers. If during the mapping process a correspondence is attempted between those agents: IN(TEA1, POT1) and IN(LIQUID, CONTAINER), then instead of building a correspondence hypothesis between them, a new agent is being built which represents a skolemized version of the general statement, namely IN(WATER1, CONTAINER1) and a correspondence hypothesis between it and IN(TEA1, POT1) will be formed. In this way the mapping process guided the process of extending the representation of the old episode, thus producing a new richer representation with inclusions, such as IN(WATER1, CONTAINER1).

In summary, AMBR dynamically forms the representation of old episodes by selecting only some of the encoded aspects of the episode (hopefully the relevant ones), and by adding new aspects which have not been explicitly encoded from beforehand – this is done either as skolemized versions of more general facts, or by borrowing facts from other episodes (blending).

The specific mechanisms proposed in AMBR for re-representation of old episodes might be psychologically valid or not, but the very fact that such dynamic re-representations are being made by humans has been shown to be valid above. Another important aspect is that this re-representation in AMBR is a result of the interplay of memory retrieval (determining which agents will be brought into WM), mapping (determining which agents are unpaired), and deductive reasoning (skolemization) and could not be realised if they were not running in parallel and interacting with each other. Finally, as I will discuss in the next section, all these complicated processes of re-representation and mapping are performed using only local information, i.e. each individual agent decides which links to establish, which new agents to form, etc.

3. FROM CENTRALIZED PLANNING TOWARDS FREE MARKET: THE NEED FOR DYNAMIC AND EMERGENT COMPUTATION

Adam Smith is not only the most famous economist who introduced the theory of the free market as a regulator of the economy and was against any form of governmental control over the market. In his book "An Inquiry into the Nature and Causes of the Wealth of Nations" (Smith, 1776) he also introduced the idea of emergent phenomena in the social sciences. He wrote about "the invisible hand by which man is led to promote an end which was not part of his intention". Thus when someone decides to start the production of certain goods in an area where the rate of profit is very high he/she does it in order to gain this high profit, however, since many will do the same, this will result in declining prices and eventually decreasing the rate of profit in this area which was in no way a goal of the producers, but they have achieved it by their actions. Von Hayek (1967), another famous economist, pro-claimed that finding an explanation of the mechanisms of these emergent phenomena is the main task of the social sciences: "those unintended patterns and regularities which we find to exist in human society and which it is the task of social theory to explain".

Some human societies were tempted to find a more direct and faster way to achieve a balance in their economy – why wait till the free market regulates prices and production when the government could calculate the desired prices and amounts of production in every economic area and directly postulate them. These attempts have recently collapsed completely. Why? The problem is that economic systems are too complex to be directly controlled and what seems to be "the more efficient direct way" is actually a very rigid way that cannot be flexible enough to reflect dynamic changes in the environment.

Cognitive scientists are gradually learning the same lesson. The attempts to build a model of human cognition based on a centralized control system are doomed to failure. No such system could be flexible enough to adapt to all dynamic changes in the environment and to reflect all possible human goals. Such a system is inherently rigid as it reflects the tasks and circumstances envisaged by its designer. An alternative approach has been proposed by Marvin Minsky (1983) which is based exactly on the analogy with human societies and has been called "the society of mind". Another alternative is the connectionist approach based on the analogy with human neural networks.

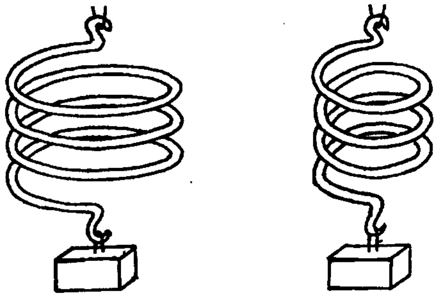


Figure 3. Illustration accompanying the target problem.

Nevertheless, we are still trying to build models of analogy-making based on the assumption that the solution of a problem is determined by its formulation and the knowledge background (including previous solutions to other problems) the subject has. Several examples of *context effects* are presented here which demonstrate that analogy-making is not that simple and predictable.

Kokinov and Yoveva (1996) conducted an experiment on problem solving where seemingly irrelevant elements of the problem solver's environment were manipulated. The material

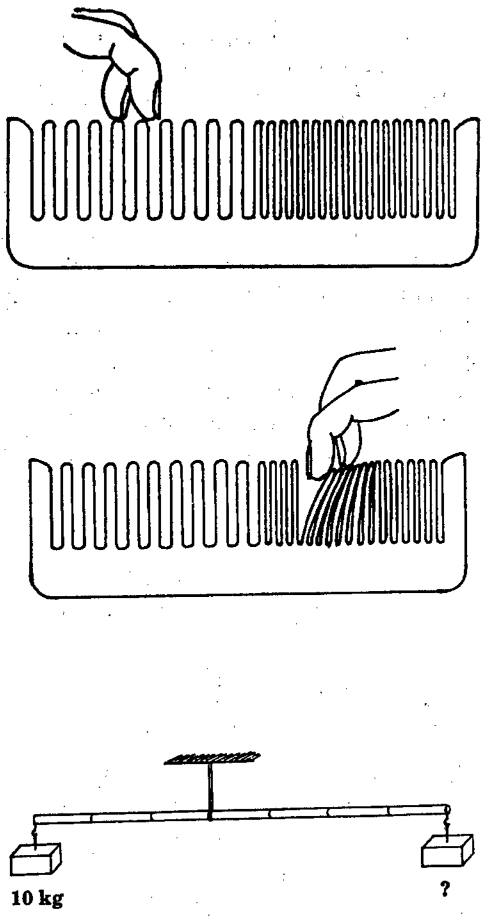


Figure 4. Illustrations accompanying the irrelevant problems in the various experimental conditions.

manipulated consisted of drawings accompanying other problems which happened to be printed on the same sheet of paper. There was no relation between the problems and the subject did not have to solve the second problem. However, these seemingly irrelevant pictures proved to play a role in the problem solving process as we obtained different results with different drawings. We used Clement's spring problem:

"Two springs are made of the same steel wire and have the same number of coils. They differ only in the diameters of the coils. Which spring would stretch further down if we hang the same weights on both of them?"

The problem description was accompanied by Figure 3 .

In different experimental conditions the drawings used as accompanying a second unrelated problem on the same sheet of paper were different: a comb, a bent comb, and a beam (Figure 4).

The results obtained in these experimental conditions differed significantly (at the 0.01 and 0.001 levels): in the control condition (no second picture on the same sheet of paper) about half of the subjects decided that the first spring will stretch more and the other half 'voted' for the second one, with only a few saying they will stretch equally. In the comb condition considerably more subjects suggested that the first spring will stretch more, in the bent comb condition considerably more subjects preferred the second spring, and in the beam condition more subjects than usual decided that both springs will stretch equally (Figure 5).

In a more recent study (the thinking-aloud experiment described in section 2) the subjects who had to solve the lightbulb problem were divided into two groups. In the control group there were no other problems on the sheet of paper, in the context group the following problem was presented on the same sheet.

"The voting results from the parliamentary elections in a faraway country have been depicted in the following pie-chart. Would it be possible for the largest and the smallest parties to form a coalition which will have more than 2/3 of the seats?"

The results are the following: in the context group *all* 7 subjects who produced the convergence solution to the lightbulb problem used *three* laser beams (7:0), while in the control group two subjects said they would use *two or three* beams and the rest said they would use either *two* or *several* beams (2:5). The difference is significant at the 0.01 level.

The results from both experiments demonstrate that sometimes small changes of a seemingly arbitrary element of the environment can radically change the outcomes of the problem solving process (can block it, or guide it into a specific direction).[1] Such phenomena are called "catastrophes". It would be very difficult to account for such effects by a model based on centralized control because in order to do so

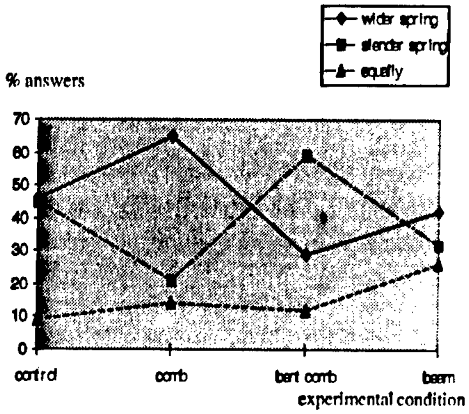


Figure 5. Percentage of proposed answer in all the experimental conditions.

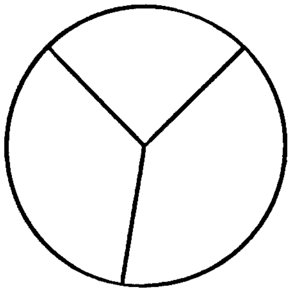


Figure 6. Illustration accompanying the context problem.

the centralized processor would have to process all possible stimuli in the perceptual field and to check whether they can be involved in the problem solving process, which would be inefficient and time-demanding to such an extent as to make it impossible.

The AMBR model adopts the following approach to accounting for context effects. It assumes that different micro-agents process different aspects of the problem and the environment. If it happens that one agent processing an arbitrary and seemingly irrelevant visual stimulus enters an interaction with a second agent processing a relevant problem aspect, then the first agent will be additionally activated and become more relevant and thus involved in the collective process of problem solving performed by the society of agents. This is a very brief and simplified description of what happens in the model, a detailed description would be based on the specific mechanisms of spreading activation, marker passing, link establishment and between-agent communication which are too complicated to be outlined in the limited space of this article.

Another important aspect of analogy-making which makes it difficult to predict whether the subject will be able to spontaneously find an analogous base (which we know he/she knows) is that this process depends on his/her preliminary internal state which is typically not related to the current problem, but is related to recently performed activities. Thus Kokinov (1990) demonstrated *priming effects* on analogical problem solving (as well as on other types of reasoning) which have a very dynamic nature, namely they are very powerful immediately after the priming event and decrease in the course of time and eventually disappear after a short period of time (in this particular study within a period of about 25 minutes). These priming effects have been qualitatively reproduced by a previous version of the AMBR model based on the pre-activation of certain agents and the decay of their activation in the course of time (Kokinov, 1994c). We plan to reproduce these priming effects with the new version by running it continuously thus solving various problems one after another.

The main conclusion from the considerations in this section is that in order to build adequate models of analogy-making, we need to base them on massively parallel architectures allowing the parallel work and interaction of many small processing entities. In addition the architecture should allow for dynamic short-term changes in the structure of interactions between these entities, something that current connectionist models do not allow.

AMBR and the underlying cognitive architecture DUAL are definitely not the best solution to these requirements. For example, top-down pressure ("the invisible hand" of the context) is limited to the current distribution of activation over agents which facilitates the local communication between agents in one direction and inhibits it in another, supports certain coalitions of agents and suppresses others. It is doubtful that this would be enough to explain all context and priming effects. On the other hand, CopyCat and TableTop have one additional top-down pressure which is called "temperature" and reflects an internal evaluation of the mental state and how close the system is to the solution of the problem. A problem with this approach is that it assumes the existence of a centralized agent watching the whole situation, computing the temperature and then communicating it back to all agents – this resembles again centralized "government" control, although it is weak control – it does not specify what the agents should do, but only changes their biases and thresholds.

The next question to be discussed in the last section is whether the mechanisms performing analogy-making can be considered domain-specific and thus form something that several researchers have called an analogy-making engine.

---

[1] This is analogous to the following phenomenon in economy – the bankruptcy of a single bank can trigger off a chain of bankruptcies and eventually a global financial crisis.

4. FROM A SPECIALISED ENGINE TOWARDS AN EMERGENT PHENOMENON: INTEGRATING ANALOGY WITH OTHER COGNITIVE PROCESSES

If analogy-making is modeled within a highly parallel architecture of "the society of mind" type, then there is no need to assume that there are mechanisms or agents which are so specific that are solely used for analogy-making. On the contrary, the analogy-making process would be considered as an emergent phenomenon, i.e. that is how we describe certain types of emergent behavior produced by the society of agents. AMBR, for example, uses mechanisms like spreading activation, marker passing, etc. which in no way may be considered as specific for analogy-making. Spreading activation, in particular, is involved in all memory processes; marker passing is involved in the processes of evaluating semantic similarity, categorization, directed search, property inheritance, etc. A process that might seem more specific for analogy-making is the ability of agents to establish hypotheses for structure correspondence (i.e. correspondence between substructures), however, this process seems so fundamental that it is doubtful that it is specifically designed for analogy-making – all processes of perception would need some structure correspondence abilities, all relational processing would also require this ability.

If we subscribe to the "emergent phenomenon" view on analogy, then it would be natural to integrate it with all other cognitive processes – simply they are emerging from the collective behavior of the same micro-agents. Then the boundaries between analogy-making, perception, memory, deductive reasoning, etc. can be described as conventional – as classification of various types of collective behavior of the same set of agents and produced by the same mechanisms (probably in different proportions). Thus Kokinov (1988, 1990, 1994c) has argued that the boundaries between analogy, deduction and generalization are a convention and that these processes are implemented by the same mechanisms. Of course, this is yet only one unsubstantiated hypothesis.

This paper is probably too general and full of speculation, however, its purpose has been neither to describe AMBR in details (which is not possible because of space limitations), nor to defend its basic principles. I am fully aware of the fact that these principles express only one possible point of view on modeling analogy-making. The purpose is to present some challenges to current models of analogy-making as seen by the author and to suggest possible ways of meeting them hoping to combine these ideas with other views expressed during the workshop.

REFERENCES

Bartlett, F. (1932). Remembering. Cambridge: Cambridge Univ. Press.

Falkenhainer, B., Forbus, K., and Gentner, D. (1986). The structure-mapping engine. *Proceedings of the Fifth Annual Conference on Artificial Intelligence*. Los Altos, CA: Morgan Kaufman.

Forbus K., Gentner D., and Law, K (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141-205.

French, R. (1995). *The subtlety of sameness: A theory and computer model of analogy-making*. Cambridge, MA: MIT Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Hofstadter, D. and the Fluid Analogies Research Group (1995). *Fluid concepts and creative analogies: Comuter models of the fundamental mechanisms of thought*. New York: Basic Books.

Holyoak K. and Koh K. (1987). Surface and structural similarity in analogical transfer. *Memory and Cognition, 15* (4), 332-340.

Holyoak K. and Thagard P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.

Holyoak, K. and Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Hummel, J. and Holyoak, K. (1997). Distributed representation of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427-466.

Keane, M., Ledgeway, K., and Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science, 18*, 387-438.

Kokinov, B. (1988). Associative memory-based reasoning: How to represent and retrieve cases. In T. O'Shea and V. Sgurev (Eds.), *Artificial intelligence III: Methodology, systems, applications*. Amsterdam: Elsevier.

Kokinov, B. (1990). Associative memory-based reasoning: Some experimental results. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kokinov, B. (1994a). The context-sensitive cognitive architecture DUAL. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale,NJ: Lawrence Erlbaum Associates.

Kokinov, B. (1994b). The DUAL cognitive architecture: A hybrid multi-agent approach. *Proceedings of the Eleventh European Conference of Artificial Intelligence*. London: John Wiley & Sons, Ltd.

Kokinov, B. (1994c). A hybrid model of reasoning by Analogy. In K. Holyoak and J. Barnden (Eds.), *Advances in Connectionist and Neural Computation Theory. Vol. 2: Analogical Connections*. Norwood, NJ: Ablex Publishing Corp.

Kokinov, B., Yoveva, M. (1996). Context Effects on Problem Solving. In: *Proceedings of the 18th Annual Conference of the Cognitive Science Society*. Erlbaum, Hillsdale, NJ.

Kokinov,B., Nikolov,V., and Petrov,A. (1996). Dynamics of emergent computation in DUAL. In A. Ramsay (Ed.), *Artificial Intelligence: Methodology, Systems, Applications*. Amsterdam: IOS Press.

Loftus, E. (1977). Shifting Human Color Memory. Memory and Cognition, vol. 5, 696-699.

Loftus, E. (1979). Eyewitness Testimony. Cambridge, MA: Harvard Univ. Press.

Minsky, M. (1986). *The society of mind.* New York: Simon and Schuster.

Mitchell, M. (1993). *Analogy-making as perception: A computer model.* Cambridge, MA: MIT Press.

Neisser, U. & Harsch, N. (1992). Phantom Flashbulbs: False Recollections of Hearing the News about the Challenger. In: Winograd, E. & Neisser, U. (eds.) Affect and Accuracy in Recall. NY: Cambridge Univ. Press

Ross, B. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 456-468.

Ross, B. and Sofka, M. (1986). [Remindings: Noticing, remembering, and using specific knowledge of earlier problems]. Unpublished manuscript.

Schunn, C. and Dunbar, K. (1996). Priming, analogy, and awareness in complex reasoning. *Memory and Cognition, 24*, 271-284.

Smith, A. (1776). Inquiry into the Nature and Causes of the Wealth of Nations.

Thagard, P., Holyoak, K., Nelson, G., and Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence, 46*, 259-310.

Von Hayek, F. (1967). Studies in Philosophy, Politics, and Economics. London: Routledge & Regan Paul.

Wharton, C., Holyoak, K., and Lange, T. (1996). Remote analogical reminding. *Memory and Cognition, 24* (5), 629-643.

Papers

TELLING JUXTAPOSITIONS: USING REPETITION AND ALIGNABLE DIFFERENCE IN DIAGRAM UNDERSTANDING

Ronald W. Ferguson, Kenneth D. Forbus
The Institute for the Learning Sciences
Northwestern University
Evanston, IL 60201 USA

ABSTRACT

Diagrams often use repetition to convey points and establish contrasts. This paper shows how MAGI, our model of repetition and symmetry detection, can model the cognitive processes humans use when reading repetition-based diagrams. MAGI, which is based on the Structure Mapping Engine, detects repetition by aligning both visual and conceptual relational structure.This lets visual regularity of form support an understanding of the conceptual regularity such forms often depict. We describe JUXTA, which uses this insight to critique a class of diagrams that juxtapose similar scenes to demonstrate physical laws.

INTRODUCTION

In explanatory diagrams, repeated structures often have special significance. To underscore a point or emphasize a difference, diagrams often juxtapose events, scenes, or objects. Examples include a "before and after" display of shirts in a laundry detergent ad and a point-by-point comparison of pumps in a physics text. In such cases,visual repetition heightens contrasts and encourages deeper comparisons. This effect is an instance of what we have termed *analogical encoding* (Ferguson, 1994), because it uses repetition and symmetry detection to support other reasoning processes.

Diagram designers have long known the utility of repetition. Edward Tufte writes that repeating structure "takes advantage of our notable capacity to compare and reason about multiple images that appear simultaneously within our eyespan. We are able to canvas, identify, reconnoiter, select,contrast, review—ways of seeing quickened and sharpened by thedirect spatial adjacency of parallel elements." (Tufte, 1997, p. 80). Repetition, detectable at a glance, aids the reader in exploring, and thus understanding, a diagram.

An example illustrates this point.Figure 1 is from a solar energy text (Buckley, 1979). This diagram illustrates a principle of heat transfer by juxtaposing two scenarios. In these scenarios, heat flows from a hot liquid, along an immersed metal bar, to amelting ice cube. Because heat flows faster in the leftmost scene, its ice cube melts more quickly. This difference between the scenarios shows how increasing a conductor's cross area increases heat transfer.

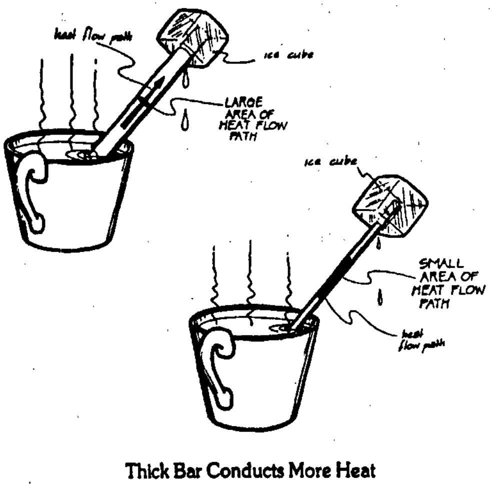

heat flow path
ice cube
LARGE AREA OF HEAT FLOW PATH
ice cube
SMALL AREA OF HEAT FLOW PATH
heat flow path
Thick Bar Conducts More Heat

Figure 1. A diagram from Buckley(1979).

The diagram uses repetition to good effect. The two scenarios not only contain the same physical elements, but are also visually similar. Before understanding the processes or the physical objects, the diagram reader may sense this visual "echo", which divides the diagram into two parts. This division signals the reader thatthese two parts are to be compared. Then, the visual correspondence of similar shapes supports the conceptual correspondence of the two cups, two bars, and two heat flows that are key to understanding the point in the caption.

If the designer had arranged the two scenes to be similar in conceptual but not visual terms—if, for example, the cup and icecube were shaped or arranged differently—the reader could still understand the diagram. But she might not instantly recognize the implicit comparison, as before. The diagram'svisual repetition allows its conceptual comparison to be quickly grasped.

This diagram is also designed so that all differences are relevant. The sole differences in the diagram are the greater thickness of the left metal bar, and the greater volume of water dripping from the left ice cube. These differences are tied to point of the caption: "Thick bar conducts more heat." The thicker bar is the independent variable, and the increased melting visibly indicates the greater heat flow.

Other differences could have been allowed. The cups could differ in volume or height, or the metal bars could differ not just in thickness, but in length. Intuitively, however, such differences would make the diagram less clear. As Tufte notes," [i]nformation consists of differences that make a difference." (1997, p. 65)

The two repetition based techniques used by this diagram—using visual regularity support a conceptual comparison, and limiting differences to only those relevant to the diagram's point—are our starting point for a cognitive model of how humans comprehend repetition in diagrams.

STRUCTURAL ALIGNMENT PROCESSES IN DIAGRAMMATIC REASONING

Why should visual repetition aid diagram comprehension? How does difference contribute to understanding? We believe the answers may lie in structure-mapping processes.

Our explanation involves two models. The first, MAGI, is a model of repetition and symmetry detection which links regularity detection with analogical mapping. The second, Markman and Gentner's *alignable difference* model, show difference detection depends on structural alignment. Based on these two models, we describe three diagram design defects that occur inrepetition-based diagrams.

MAGI

Similarity and analogical comparison can be modeled as the structural alignment of propositional descriptions.(Falkenhainer, Forbus, & Gentner, 1989; Forbus,Ferguson, & Gentner, 1994; Gentner, 1983; Gentner, 1989; Goldstone, 1994;Holyoak & Thagard, 1989; Keane & Brayshaw, 1988).

MAGI (Ferguson, 1994, In preparation) isthe first model linking regularity detection with similarity. MAGI is basedon the idea that symmetry and repetition (both visual and conceptual) can be viewed as asimilarity mapping between a description and itself. Using an extension of the Structure Mapping Engine (SME; MAGI uses structural alignment to detect regularity within a single description. Like SME, MAGI's mapping process is computationally tractable because it operates in a local-to-global fashion.Individual alignments are constructed in parallel and then aggregated into global mappings, mappings governed by systematicity constraints favoring relationally deep, interconnected correspondence sets. MAGI also operates incrementally. As new information is added to a description, MAGI's mapping can be extended appropriately.

To detect regularity, MAGI maps over a visual representation built by Geo Rep. Geo Rep, given a line drawing, builds a propositional description of its salient perceptual relations. Starting with the drawing's graphical primitives (line segments, arcs, circles, ellipses, and spline curves), a set of visual routines (Ullman, 1984) represent a variety ofrelationships, including types of object connection, parallelisms,horizontal and vertical relations, and descriptions of polygons and their inflexion points. Geo Rep contains a rule engine, and its default rule set can be extended to handleparticular domains.

Given a stylized line drawing of our example diagram (Figure 2), MAGI can map over the diagram's perceptual relations to determine object correspondences figure 3. If we add information about the physical objects and processesin the diagram, MAGI can extend its mapping accordingly.

MAGI canhelps explain the immediacy and utility of visual regularity. It describe show repetition is detected and the nature of the correspondences produced. More importantly, however, it provides a link between perceptual and conceptual regularity.

Based on MAGI's model, we assume the reader of the diagram begins by detecting the its visual regularity (Figure 3). As conceptual information is also acquired, the reader may attempt to use this information to extend the mapping. However, if the new conceptual information cannot be mapped consistently with thevisual information, the reader may either fail to notice the conceptual regularity, or need to ignore the previous visual regularity. Handling this conflict may slow or blockdiagram comprehension.

Visual repetition and symmetry detection operate very early inperception. Visual symmetry can be detected after display times of lessthan 100 ms. (Carmody, Nodine, & Locher, 1977; Corballis &Roldan, 1975; Julesz, 1971). Consequently, most models of symmetry detection do not incorporate more complex algorithms such asstructural alignment (with the notable exception of the Wageman's Bootstrapping model (1995)). Until recently, it seemed unlikely that visual symmetry detection couldinvolve alignment.

However, new results from Aminoff, Ferguson and Gentner(In preparation; 1996) provide evidence that even the earliest forms of symmetry detection may involve alignment. In two experiments, Aminoff *et al.* (in preparation) showed subjects symmetric and asymmetric polygons with display times of 50 ms. In each experiment, subjects were consistently faster or more accurate at judging the asymmetry of polygons contain ingaligned qualitative differences,including differences in corner concavity and number of vertices. This effect was independent of several other quantitative asymmetry measures,including differences in area and radial length. Thus, these results are new evidence for alignment early in symmetry detection. For this reason,it is entirely possible that structural alignment is used for both very early and much later forms of regularity detection.

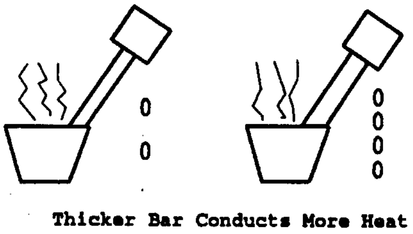


Figure2. Stylized redrawingof Figure .

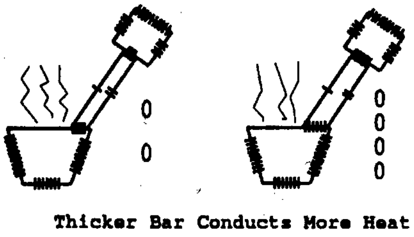


Figure 3. Regularity found by MAGI astructural alignment.

ALIGNABLE DIFFERENCES IN COMPARISON

The MAGI model explains how visual repetition can support an understanding of conceptual repetition. However, we have not yet addressed the utility of differences in a diagram.

Of course, difference detection might be seen as a very different process than repetition detection. In Tversky's influential contrast model of comparison (1977), similarity increases as a function of common features, and decreases as afunction of mismatched features. If individual features are assumed independent, the detection of matched features would neither encourage nor block the detection ofmismatched features.

Studies by Markman and Gentner, however, found evidence that alignment significantly affected the kinds of differences human participants noticed, with most differences directly linked to preexisting aligned commonalities (and thus called *alignable differences*). This model predicts that increasing the similarity of two concepts also increases the number of alignable differences noticed. This prediction was borne out in their experiments. When human participants were asked to list differences between high and low similarity word pairs (Markman & Gentner, 1993),participants consistently listed more alignable differences for pairs with high similarity (hotels and motels) than for pairs with low similarity(magazine and kitten). A second set of experiments (Markman & Gentner, 1996),generalized the results for word pairs to pairs of pictures, and also showed that alignable differences had a greater effect onparticipants' judgment of similarity than did nonalignable differences. When determining differences between twothings, people seem to focus more on alignable than non alignable differences.

Because alignable differences are produced more often than nonalignable differences, and because they have a greater influence on participants' similarity judgments, it is safe to assume that alignable differences are critical to the contrasts undertaken in repetition-based diagrams. Because alignable differences are easily generated in the context of structural alignment,visual alignable differences may communicate their points very effectively.

We conjecture that structural alignment has a profound effect ondiagram understanding. Visual alignment supports conceptual alignment, andalso highlights alignable differences.

THREE PROBLEMS OF DIAGRAM STYLE

Which factors—by analogy with understanding writtenprose—make a diagram more comprehensible? As we have seen,repetition in diagrams should be visually apparent, and should draw the reader into a deeper conceptual comparison without causing missteps or misalignments.Alignable differences should be salient and should serve the point of the diagram. These criteria suggest three general types of design defects that may hinder comprehension ofrepetition-based diagrams.

Visual/conceptual cross-mappings. Cross-mappings(Gentner & Toupin, 1986) occur whensurface information and relational information suggest different mappings for the same objects. Visual cross-mappings occur when two objects are visually alignable, but the roles or functions of the aligned objects are not equivalent. For example, if two oblong objectsmatch, but one is a metal bar conducting heat, and another the handle of acontainer, the initial visual correspondence between the parts mightconfuse readers. The readers might seek some common functional role between the two objects, and find none, slowing them down.

Alignable differences that are either not salient or not compelling. Some alignable differences are more noticeable than others.In our example diagram, for instance, many people find the difference in the number of water droplets easier to spot than the difference in thickness for the two metal bars.

We do not yet have a theory of what makes alignable differences salient or compelling. Understanding salience alone requires a more complex model of visual attention than we have available. However, we can define techniques to make alignable differences either more salient or more compelling, a process we call *difference amplification*.

We can make alignable differences more salient by either adding additional alignable structure that draws attention to that difference,or by other techniques, such as color. Besides making differences more salient, we can also make them more compelling by making the importance of the difference more evident. We do this by making it easier for the diagram reader to link the visual alignable differences to the conceptual differences underlying the diagram's point. Labeling is the easiest way to accomplish this.

Aligned differences unrelated to, or interfering with, the pointof diagram. When alignable differences exist, they should be relatedto the diagram's point. Some alignable differences may be irrelevant; if our diagram had one cup colored red, and the other blue, this difference would be obvious but unlikely to confuse the reader. Alignable differences may detract from a diagram when they appear to be related to the point of the diagram, but are not. If one cup was being heated with a burner in our diagram, we might be confused about how this particular difference relates to the role of the thicker metal bar, since both the flame and relative bar thickness would affect the rate of heat flow.[1] Such alignable differences make it more difficult to draw a conclusion from the diagram, and thus hinder the reader's ability to comprehend the point of the juxtaposed situations.

To summarize, the MAGI model and Markman and Gentner'salignable difference model suggest three ways in which a diagram can beconfusing. First, it may contain a visual-conceptual cross-mapping. Second, alignable differences may not be salient.Finally, alignable differences may be irrelevant or may interfere with the point of the diagram. These three criteria can be easily characterized in terms of the MAGI model and some simple assumptions about visual representation.

Because these stylistic problems can be cleanly described in terms of the MAGI model, it is possible to build a diagram critic that uses these principles to parse and critique diagrams. We can use mismatches between correspondences at the visual, physical and process levels to determine how well the visual regularity in the figure guides the comparison. If we have a representation of the diagram's point (which often can be derived from the caption), we can also determine if the alignable differences in the figure convey the point, are orthogonal to the point, or get in the way of understanding the point.

We have built such a system, called JUXTA[2]. Given diagrams that juxtapose physical situations, JUXTA can produce a critique of the figure, and note differences that may confuse ordistract the reader. JUXTA also amplifies a diagram's relevantalignable differences by labeling them, using its physical knowledge tocreate and place useful explanatory labels.

We now summarize how JUXTA works.

THE *JUXTA* ARCHITECTURE

Figure 4 describes JUXTA's architecture. JUXTA's inputis a stylized line drawn diagram and a representation of the diagram's caption. It provides three kindsof feedback. First, it amplifies relevant alignable differences bylabeling them with process descriptions. Second, it critiques differencesthat interfere with the point of the diagram (as given in the caption). Finally, it notes differences thatare orthogonal to the point of the diagram, and thus may be removed at thedesigner's discretion.

---

[1] Tufte (1997) gives an example of how this principle is violatedin the "before and after" drawings done by the 19th century architect Humphrey Repton. Repton's "after"views often embellish. For example, a landscaping proposal adds changes to the "after" view that are appealing but are unrelated to the proposed modification,such as stylishly dressed people on the sidewalks and fine sailing ships inthe adjacent harbor.

[2] JUXTA stands for Juxtaposition Understanding and Explanation Through Analogy.

Processing the figure

First, JUXTA (using the GeoRep visual representation engine) represents the diagram at three different levels—visual level (e.g. a square), a physical level (an ice cube), and aphysical process level (heat flowing into an ice cube) using a set of rulesand low-level visual routines (Figure 5).

JUXTA uses a simplified model of object recognition, which depends on a set of rules to determine when a set of visual entities represent a particular type of structured object. The heuristics used for object recognition are summarized inTable1. This technique requires the use of stylized diagrams, but otherwise retains much of the flexibility of general diagrams. For example, objects can be drawn using a drawing program, object dimensions can vary as needed, and diagram parts can be composed into more comprehensive scenes

Of course, JUXTA also needs a representation of the caption, which isassumed to contain the point of the diagram. To avoid doing natural language interpretation, we give JUXTA the representation of the caption directly. The representations use Qualitative Process Theory (Forbus, 1984). It is useful to identify two parts of captions for juxtaposition diagrams, the *antecedent* and *consequent*. In this caption, the antecedent is the difference in thickness of the bars and the consequent is the difference in the rates of heat flow.

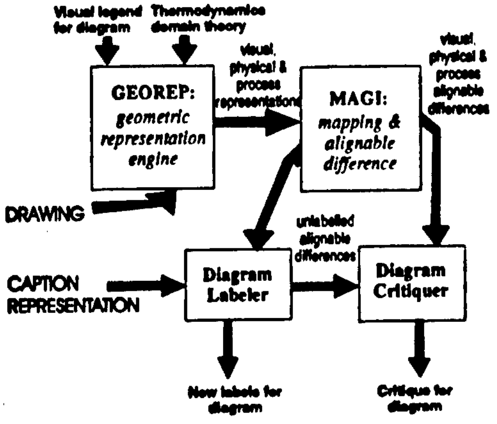

Visual legend for diagram
Thermodynamics domain theory
GEOREP: geometric representation engine
visual, physical & process representations
MAGI: mapping & alignable difference
visual, physical & process alignable differences
DRAWING
unlabelled alignable differences
CAPTION REPRESENTATION
Diagram Labeler
Diagram Critiquer
New labels for diagram
Critique for diagram

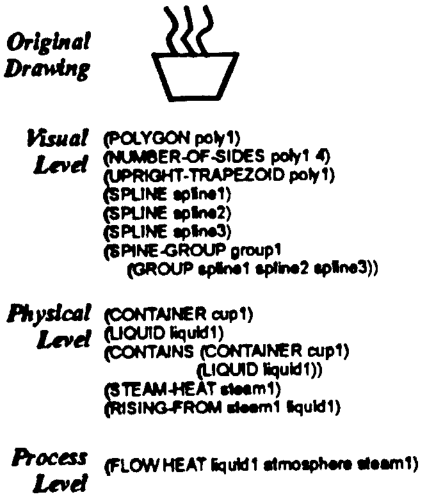
Figure4. JUXTA's architecture.

Finding regularity and differences

JUXTA runs MAGI on the figure to detect correspondences (Figure 3).JUXTA then uses a simple mechanism for detecting alignable differences based on finding differences in dimensions predetermined by the object category.For example, when two trapezoids correspond, JUXTA compares their height and length. Invisible differences, such as differences in the rate of aphysical process, are inferred from visual differences via rules in adomain-dependent knowledge base. For example, if the two trapezoids represent two cups, and one trapezoid is larger, then JUXTA infers that the cup represented by that trapezoidhas greater volume. This way, visible differences enable JUXTA to infer deeper conceptual differences.

Amplifying differences via labeling

At this point, JUXTA now has analyzed the figure at the visual, physical,and process levels. It also has, for each of those levels, computed the representation of that level, the regularity mapping for that level, and the set of


Original Drawing
Visual Level
Physical Level
Process Level

```
(POLYGON poly1)
(NUMBER-OF-SIDES poly1 4)
(UPRIGHT-TRAPEZOID poly1)
(SPLINE spline1)
(SPLINE spline2)
(SPLINE spline3)
(SPINE-GROUP group1
   (GROUP spline1 spline2 spline3))

(CONTAINER cup1)
(LIQUID liquid1)
(CONTAINS (CONTAINER cup1)
            (LIQUID liquid1))
(STEAM-HEAT steam1)
(RISING-FROM steam1 liquid1)

(FLOW HEAT liquid1 atmosphere steam1)
```

Figure5. Levels of representation.

rtext170 **Class of object**	**dxfrtext170** **Visuallegend**	**w3941** **Salient dimensions**
rtext170***Container of liquid***	**1**Upright,top-heavy trapezoid	**rtext170**Height andwidth
rtext170***Steam or heat***	**1**Group ofproximate spline curves	**rtext170** Number ofcurves
rtext170***Metal bar***	**1**Oblongoblique trapezoid	**rtext170**Length andthickness
rtext170***Ice cube***	**1**Square	**rtext170**Width
rtext170***Water drops***	**1**Group ofproximate, vertical ellipses	**rtext170** Number ofellipses

Table1: Visual legend forrecognized objects

Table1. Visual legend forrecognized objects.

aligned differences.It can now begin its critical analysis of the figure. First, JUXTA attempts to link the aligned differences to the antecedents and consequences of the point given in the caption. It then amplifies the aligned differences by labeling them. The labels link each key difference in the caption to some visual difference.

To link the objects in the diagram with the referents in the caption representation, JUXTA matches the caption representation against the physical and process representations of the diagram, and uses this match to fill the caption representation's unfilled slots. This is how, for example, JUXTA figures out which objector objects the caption's "thicker bar" refers to.In this case, JUXTA can find the thicker bar on the right using the common object category (metal bar) to select both metal bars, and using thealignable difference (thicker) to distinguish between them.

Once JUXTA understands which parts of the figure are being referenced in the caption, it labels the differences. This involves constructingpaired labels for each alignable difference given in the caption, and then determining where to place each label.To label an alignable difference, JUXTA must find a visible referent to point to. When an alignable difference isalong a visible dimension (such as the thickness of a bar), the object itself is the referent of the label, and JUXTA points to the shape which represents the physical object. Alternatively, when a caption relationship is not visible (such as heat flowalong the metal bar), JUXTA looks for a consequence of the relationship which is visible difference. In the example figure, the difference in heatflow causes a difference in the rate at which the ice cube melts, causing a visible difference in the number of drops (ellipses), so JUXTA labels this. The result of the labeling stage on the example diagram is given in Figure 6.

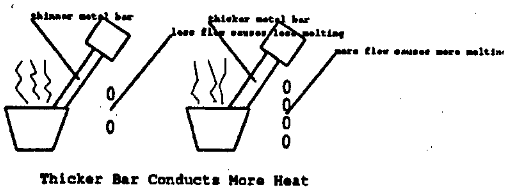

thinner metal bar
thicker metal bar
less flow causes less melting
more flow causes more meltin
Thicker Bar Conducts More Heat

Figure 6. Results of labeling stage of JUXTA on examplediagram.

Critiquing the diagram

After labeling the figure, JUXTA critiqueshow well the alignable differences contribute to the point of the caption.To do this, JUXTA looks at all alignable differences left over from the labeling stage. These are differences that arenot related to alignable differences referenced in the caption. If are maining alignable difference is not the result of the caption antecedent,but can have an effect on its consequent, JUXTA notes it as potentially confusing. For example, Figure 7 is a variant of our example diagram that contains this problem. Here, the amount of heat rising from the second cup is larger than the first container. JUXTA notes this difference as confusing because the amount of heat from the container implies that the second container may contain a hotter liquid, which would also increase the heat flow rate.

Of course, remaining alignable differences may not relate to the caption at all. In this case, JUXTA will not mark it as confusing, but will note the orthogonal status of the alignable difference. For example,in Figure 7, JUXTA will note that one spline curve in the leftmost group is longer. Removing this differences might make interpretation somewhat simpler, but it will not cause problems if left unchanged.

CONCLUSION

Analogical encoding techniques, based on current models of analogy and similarity, can provide key insights into diagrammatic reasoning. We have shown how MAGI, which uses structure mapping to detect repetition and symmetry, may explain how visual and conceptual regularity support one another, and how alignable differences emphasize relevant points. This model is strong enough to build a system, JUXTA,that can parse, analyze, and critique a diagram by analyzing how correspondences and differences interact between the visual, physical andprocess levels.

While JUXTA demonstrates the basic principles behind a whole class of diagrammatic reasoners, the current implementation is limited. JUXTA has only been used on a handful of figures. The recognition of objects and processes remains brittle.We are exploring similarity-based feature re-interpretation as onemechanism for improving the system's flexibility.

JUXTA also deals solely with diagrams that use binary repetition to demonstrate physical laws. In practice, diagrams use many types ofregularity, including matrices, multiple repeated items, sequences, and symmetry. To expand the kinds of regularity JUXTA handles, MAGI itself may need to be extended to handlesome forms of n-ary symmetry and repetition. This problem relates more topsychology than programming—althoughit is relatively simple to configure a version of MAGI that recognizessmall multiples of a scene, it does not yet do so in an efficient way, nordoes it reflect our understanding of how humans recognize other forms of regularity. We expect, however,that JUXTA soon handle symmetry as well as repeating diagrams.

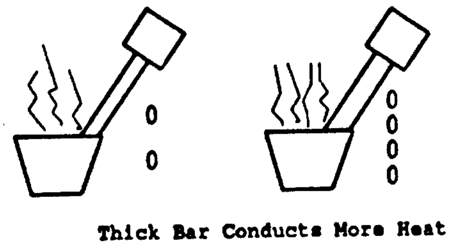


Figure 7. A faultyvariant of Figure 1.

The just-mentioned variety of diagrammatic regularity speaks to the fascinating richness of this particular sub-area of cognition. If thesimple mechanisms of JUXTA can be extended to a larger range of diagrams,they may not only provide a foundation for computer systems that can understand diagrams in amore human-like fashion, but may also have interesting consequences for our understanding of diagrammatic reasoning, regularity, and analogy.

ACKNOWLEDGEMENTS

Thanks to Dedre Gentner and Alex Aminoff for many helpful comments on this research. This research was supported by grants from the Office of Naval Research and DARPA.

REFERENCES

Aminoff, A., Ferguson, R. W., & Gentner, D. (In preparation). Early detection of qualitative symmetry.

Buckley, S. (1979). *Sun Up to Sun Down*. New York: McGraw-Hill.

Carmody, D. P., Nodine, C. F., & Locher, P. J. (1977). Global detection of symmetry. *Perceptual and Motor Skills, 45*, 1267-1273.

Corballis, M. C., & Roldan, C. E. (1975). Detection of symmetry as afunction of angular orientation. *Journal ofExperimental Psychology:Human Perception and Performance, 1*, 221-230.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence, 41*, 1-63.

Ferguson, R. W. (1994). MAGI: Analogy-based encoding using symmetry and regularity. In *Proceedings of the 16th Annual Conference ofthe Cognitive Science Society* (pp. 283-288). Atlanta: Erlbaum.

Ferguson, R. W. (In preparation). MAGI: A model of symmetry and repetition detection. .

Ferguson, R. W., Aminoff, A., & Gentner, D. (1996). Modeling qualitative differences in symmetry judgments. In *Proceedings of the18th Annual Conference of the Cognitive Science Society*. Hillsdale: Erlbaum.

Forbus, K. D. (1984). Qualitative Process Theory. *Artificial Intelligence, 24*, 85-168.

Forbus, K. D., Ferguson, R. W., & Gentner, D. (1994). Incremental structure mapping. In *Proceedings of the 16th Annual Conference ofthe Cognitive Science Society* (pp. 313-318). Atlanta, GA: Lawrence Erlbaum Associates.

Gentner, D. (1983). Structure-Mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Gentner, D. (1989). The mechanisms of analogical learning. In S.Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning*(pp. 199-241). London: Cambridge University Press.

Goldstone, R. L. (1994). Similarity, interactive activation, andmapping. *Journal of Experimental Psychology: Memory and Cognition, 20*, 3-28.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.

Julesz, B. (1971). *Foundations of Cyclopean Perception.*University of Chicago Press.

Keane, M., & Brayshaw, M. (1988). The Incremental Analogy Machine. In *Proceedings of the Third European Working Session on Learning* (pp.53-62). London: Pitman.

Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32*(4), 517-535.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparison. *Memory and Cognition,24*, 235-249.

Tufte, E. R. (1997). *Visual Explanations*. Cheshire, CT:Graphics Press.

Tversky, A. (1977). Features of similarity. *Psychological Review,84*(4), 327-352.

Ullman, S. (1984). Visual routines. In S. Pinker (Ed.), *Visual Cognition* (pp. 97-159). Cambridge: MIT Press.

Wagemans, J. (1995). Detection of visual symmetries. *Spatial Vision, 9*(1), 9-32.

MAKING SENSE OF ANALOGIES IN METACAT

James B. Marshall, Douglas R. Hofstadter
Center for Research on Concepts and Cognition
Indiana University
Bloomington, Indiana USA 47505
{marshall,dughof}@cogsci.indiana.edu

ABSTRACT

This paper outlines the main ideas and objectives of the Metacat project, an extension of the Copycat computer model of analogy-making and high-level perception. The principal features of Metacat that allow it to make sense of analogies suggested to it by the user are described using a simple example.

INTRODUCTION

The Copycat computer model of analogy-making and high-level perception was originally developed by Hofstadter & Mitchell as a computational model of subcognitive mechanisms underlying human cognition, in which the notion of *fluid concepts* plays a central role. Copycat models the process of analogy-making within a stripped-down microworld of tiny, idealized situations represented as short strings of letters. For example, a typical Copycat problem is the following: "If **abc** changes to **abd**, how does **mrrjjj** change in an analogous way?" This microworld, though austere, harbors a surprisingly rich variety of subtle problems in which a wide range of answers is almost always possible—often including deeply elegant but non-obvious ones. For example, there are many defensible answers to the above problem, including **mrrkkk**, **mrrjjk**, **mrrjjd**, **mrrddd**, **mrrjjj** (in which only **c**'s are seen as changing), **mrsjjj**, **mrdjjj**, **mrrjjjj**, **mrrkkkk**, or even **abd** or **abbddd**. The apparent simplicity of Copycat's domain is deceptive, for it remains a formidable challenge to develop a computational model exhibiting a level of creative and flexible behavior comparable to that of humans even in this tiny, restricted domain of letter-strings.

Copycat discovers analogies between different situations by building up an understanding of the situations in terms of concepts that it understands about the letter-string world. Representations of these concepts are hard-wired into the program, yet they are not static entities with sharply defined boundaries. Rather, their boundaries are inherently fuzzy, overlapping each other to varying degrees and changing in response to competing contextual pressures that arise during the course of processing. The dynamic, "fluid" nature of Copycat's concepts is intended to model the extremely flexible human ability to perceive dissimilar things as being in fact "the same" when viewed at some appropriate level of description.

A detailed exposition of the Copycat program can be found in [Mitchell, 1993] and [Hofstadter and FARG, 1995]. In this paper, we give just a brief summary of Copycat and then discuss in more detail recent work aimed at extending the model. The goal of the current project, dubbed Metacat, is to increase the program's "awareness" of its own behavior as it solves analogy problems, so that it may gain deeper insights into the analogies it makes.

THE COPYCAT MODEL

When Copycat is given an analogy problem to work on, it starts out with the letter-strings in its *Workspace*, the architectural component of the program in which all perceptual processing occurs. Small, nondeterministic processing agents called *codelets* notice relations among the individual letters and build new structures around them, organizing them into a coherent high-level picture. All processing occurs through the collective actions of many codelets working in parallel, at different speeds, on different aspects of an analogy problem, without any centralized executive process controlling the course of events. The stochastic behavior of codelets is dynamically biased by the time-varying pattern of activation in the program's network of concepts, called the *Slipnet*, that it uses to build up an understanding of an analogy problem. In turn, this context-dependent pattern of conceptual activity in the Slipnet is itself an emergent consequence of codelet processing in the Workspace.

For example, in order to discover an answer to the problem **"abc => abd; mrrjjj => ?"**, codelets work together to build up a strong, coherent mapping between the *initial* string **abc** and the *target* string **mrrjjj**, and also between the initial string and the *modified* string **abd**. Within each letter-string, codelets attempt to build hierarchical *groups*, effectively organizing the strings (the raw perceptual data) into coherent, chunked wholes. In **mrrjjj**, for example, codelets might build the "sameness-groups" **rr** and **jjj**, causing the *sameness-group* concept in the Slipnet to become activated, which in turn makes it more likely for the program to regard **m** as a sameness-group of length one within the context of the other groups in its string. A higher-level "successor-group" comprised of **m**, **rr**, and **jjj** encompassing the entire string can then be seen based on the concept of *group-length* (*i.e.*, *1–2–3*) rather than on *letter-category*. Consequently, the letter-category-based successor-group **abc** can be mapped as a whole onto the length-based successor-group **mrrjjj**, representing the recognition of these strings as instances of the same concept, even though their surface resemblance is negligible. The distributed nature of codelet processing interleaves the chunking process with the mapping process, and as a result, each process influences and drives the other.

A mapping consists of a set of *bridges* between corresponding letters or groups that play respectively similar roles in different strings. Each bridge is supported by a set of *concept-mappings* that together provide justification for perceiving the objects connected by the bridge as corresponding to one another. For example, a bridge might be built between **c** in **abc** and **jjj** in **mrrjjj**, supported by the concept-mappings *rightmost => rightmost* and *letter => group*, representing the idea that both objects are rightmost in their strings, and that one is a letter and the other a group. Non-identity concept-mappings such as *letter => group* are called *slippages*, and form the basis of Copycat's ability to perceive superficially-dissimilar situations as being identical at a deeper level.

Once a strong, coherent mapping has been built between the initial string and the modified string, another type of structure, called a *rule*, may get created based on this mapping, which succinctly describes the way in which the initial string changes into the modified string. There are often several possible ways of describing this change, some more abstract than others. For example, two possible rules for **abc => abd** are *Change letter-category of rightmost letter to successor* and *Change letter-category of rightmost letter to d.*

Different ways of looking at the initial/modified change, combined with different ways of building the initial/target mapping, give rise to different answers. The configuration of structures in the Workspace collectively represents the way in which a given analogy problem is interpreted. A particular interpretation implies a particular answer for the problem. To produce an answer, the rule describing the way the initial string changes is translated into a new rule that applies to the target string, based on the slippages underlying the initial/target mapping. For example, if the **abc => abd** change is described according to the first rule above, and the abstract successor-group similarity between **abc** and **mrrjjj** has been noticed, then the rule will be translated as *Change length of rightmost group to successor*, yielding the answer **mrrjjjj**. On the other hand, if this deep similarity has not been noticed, the answers **mrrkkk**, **mrrjjk**, **mrrddd**, or **mrrkkd** may be found instead, depending on the rule chosen and whether or not **c** in **abc** is seen as corresponding to the **jjj** group or to just the rightmost letter **j** in **mrrjjj**.

As this example suggests, Copycat's stochastic processing mechanisms enable it to find a range of different answers for a given analogy problem. Copycat attaches a rough numerical measure of "quality" to the answers it finds, which, for many problems, corresponds reasonably well to human judgments of relative answer quality. But the program has very little awareness of how it actually finds the answers that it finds. It has almost no insight into its own processing mechanisms—fluid and flexible though they may be—which guide it through the "space" of possible interpretations of an analogy problem. This is not too surprising, however, given that Copycat was intended primarily as a model of subcognitive mechanisms. All of the nondeterministic codelet activity occurring in the Workspace—the building of bridges and groups, the making of slippages, and so on—is intended to represent perceptual activity carried out below the level of "conscious awareness". In contrast, the focus of Metacat is on developing mechanisms that support a higher "cognitive" level on top of Copycat's subcognitive level. To do this, Metacat needs to be able to remember what happens while its subcognitive mechanisms are building, destroying, and reconfiguring Workspace structures in pursuit of an answer to the problem at hand, and to build explicit representations of this activity.

METACAT'S OBJECTIVES

Hofstadter has outlined several important objectives for the Metacat project [Hofstadter and FARG, 1995, Chapter 7]. First of all, the program should be able to explicitly characterize the *essence* of an answer—the core idea or cluster of ideas underlying the answer that fundamentally distinguishes it from other possible answers. The ability to perceive what a given answer is really "about" should enable the program to give at least a limited explanation of the answer's strengths and weaknesses compared to other answers it may have previously found. For example, the essence of the **mrrjjjj** answer described earlier lies in seeing both **abc** and **mrrjjj** as successor-groups, one based on the concept of *letter-category* and the other based on the concept of *group-length*. The recognition of this abstract similarity between the strings is what fundamentally distinguishes the answer **mrrjjjj** from other, more straightforward answers such as **mrrkkk**, **mrrjjk**, or **mrrddd**, in which the hidden "successorship fabric" of **mrrjjj** remains unnoticed.

The ability to compare and contrast answers, however, implies the ability to remember more than one at a time. In Copycat, answers are not retained after they are found. When Copycat discovers an answer to a problem, it simply reports the answer, along with the answer's numerical measure of quality, and then stops. No recollection of previously found answers is possible on subsequent runs of the program, so there is no way for the program to bring its past experience to bear on its current situation. This makes comparison of different answers impossible, either within a single analogy problem or across different problems. In contrast, Metacat should remember the answers it finds, along with characterizations of the key ideas involved, gradually building up in its memory a repertoire of experience on which it can draw when confronted with new situations.

In addition to remembering the answers it finds, Metacat should also keep track of patterns that occur in its own processing while it is trying to discover new answers. As it works on an analogy problem, it should create an explicit sequential trace of its own behavior as it searches through the space of possible interpretations leading to different answers. This type of memory is of a more short-term, temporal nature than that just described for the answers themselves. Such a *self-watching* ability would enable Metacat not only to remember the important events that led it to find an answer, but also to recognize when it has fallen into a repetitive or otherwise unproductive pattern of behavior. Recognizing that it is in a "rut" should enable it to subsequently "jump out of the system" by explicitly focusing on ideas other than the ones that seem to be leading it nowhere. This type of self-awareness pervades human cognition. People can easily pay attention to patterns in their own thinking; see for example [Chi et al., 1989, VanLehn et al., 1992].

Once Metacat has the ability to size up the answers it finds in terms of their essential features, it ought to be able to evaluate other answers suggested to it by some outside agent. In other words, Metacat should not only be able to come up with answers to analogy problems on its own, it should also be able to justify answers on their own terms, even if the program itself didn't come up with them. This constitutes an ability to work "backwards" from a given answer toward an insightful characterization of the answer, in order to understand why it makes sense. Once an answer has been understood in this way, it could then be compared and contrasted with other answers that the program has either itself discovered previously, or been shown by someone else.

THE METACAT MODEL

The Metacat architecture includes all of Copycat's main architectural components, such as the Workspace, the Slipnet, and the mechanisms that support distributed, nondeterministic codelet processing. In addition, new architectural components have been incorporated into the model, and mechanisms for building bridges and creating rules have been extended and generalized. These components provide a general framework in which to address the objectives outlined in the previous section.

Unlike Copycat, Metacat incorporates a memory for its answers, which allows it to remember more than one answer over the course of a run. Whenever it finds a new answer, instead of simply stopping, Metacat pauses to display the answer along with the Workspace structures representing the interpretation of the problem. This information is packaged together and stored in Metacat's memory, after which the program continues searching for alternative answers to the problem. Gradually over time, a series of answers accumulates in memory, each one representing a different way of making sense of the analogy problem at hand.

The most important type of auxiliary information stored with answers consists of structures called *themes*. Themes reside in Metacat's *Themespace*, and represent key concepts underlying the mappings created between letter-strings. Collections of themes serve as high-level characterizations of Metacat's answers, and provide a basis on which to compare and contrast answers with each other. Themes are comprised of Slipnet concepts, and assume time-varying levels of activation ranging from –100 to +100, depending on the extent to which the ideas they represent are present or absent in a particular configuration of Workspace structures.

Unlike Copycat, Metacat allows the user to suggest a particular answer to a given analogy problem. The program then tries to find an interpretation of the problem that leads to the answer in question. As an example, consider the problem **"abc => abd; xyz => ?"** with the answer **wyz** suggested to the program by the user. When run on this problem, Metacat attempts to justify the **wyz** answer by searching for an overall interpretation of the problem in which this particular answer makes sense. After several hundred codelets have been run, structures built in the Workspace typically include horizontal bridges comprising the **abc => abd** and **xyz => wyz** mappings in whicheach string is seen as mapping onto its counterpart in a straightforward, left-to-right way (*i.e.*, **a–a**, **b–b**, **c–d**, **x–w**, **y–y**, and **z–z** bridges). Also, vertical bridges map **abc** and **xyz** onto each other in a similarly straightforward, left-to-right way (*i.e.*, **a–x**, **b–y**, and

c–z). In addition, the rule *Change letter-category of rightmost letter to successor*, describing how **abc** changes to yield **abd**, and the rule *Change letter-category of leftmost letter to predecessor*, describing how **xyz** changes to yield **wyz**, both get created.

Several themes in the Themespace get activated in response to the creation of these various Workspace structures. Specifically, four horizontal-bridge themes characterizing the horizontal **abc** => **abd** bridges become activated to different degrees. Two of these themes represent the ideas of letter-category sameness and letter-category successorship within the **abc** => **abd** mapping. The **a–a** and **b–b** bridges both involve the idea of letter-category sameness, while the **c–d** bridge involves the idea of successorship. Therefore, the themes *Letter-Category:Sameness* and *Letter-Category: Successor* are both active in the Themespace, although the successorship theme is not as active as the sameness theme.

On the other hand, all bridges map objects of identical string-position (*e.g., leftmost* => *leftmost*) and object-type (*e.g., letter* => *letter*) onto each other, so the themes *String-Position:Sameness* and *Object-Type:Sameness* are highly active. These themes together serve as an abstract characterization of the **abc** => **abd** mapping. Other sets of themes in the Themespace characterize other Workspace structures in a similar fashion.

Thus, themes are first and foremost representational structures. But under certain conditions, when highly activated, they can also exert powerful *top-down pressure* on Metacat's processing mechanisms, strongly biasing the stochastic behavior of codelets in favor of particular outcomes. Active themes can be regarded as Metacat's way of "seizing on" certain key ideas implicit in an analogy problem and making them explicit, driving the program toward an interpretation of the problem organized around these ideas.

In the above example, Metacat perceives **abc** and **xyz** as successor-groups going in the same direction (left-to-right). This is represented by the vertical **a–x** and **c–z** bridges, which are supported by the concept-mappings *leftmost* => *leftmost* and *rightmost* => *rightmost*, respectively. However, this way of interpreting the situation doesn't make sense, because **c** and **x** are not seen as corresponding to each other (since there is no bridge between them), yet they are both identified by the rules as being the objects that change in their respective strings (the **c** to its successor and the **x** to its predecessor).

At some point, codelets may compare the two rules and notice that taken together, they imply the concept-mappings *rightmost* => *leftmost* and *successor* => *predecessor*. These concept-mappings suggest the idea of mapping the strings **abc** and **xyz** onto each other in a *crosswise* fashion, so that one group is viewed as a successor-group and the other is viewed as a predecessor-group, with the rightmost letter of one corresponding to the leftmost letter of the other, and vice versa. This idea can be succinctly characterized by a set of vertical-bridge themes representing string-position and group-direction *oppositeness*. These themes are clamped by codelets at full activation, strongly promoting the creation of new structures compatible with the idea of a vertical crosswise mapping and greatly weakening existing structures incompatible with this idea.

For example, the **a–x** and **c–z** bridges are incompatible with the idea of mapping **abc** and **xyz** onto each other in opposite directions, represented by the *String-Position:Opposite* theme, since they are supported by *leftmost* => *leftmost* or *rightmost* => *rightmost* concept-mappings. They are thus easily broken and replaced by **a–z** and **c–x** bridges, which are compatible with this idea. The net effect is that the original vertical mapping described above is swiftly reorganized by codelets into a new mapping consistent with the activated themes.

Eventually, the burst of new structure-building activity caused by clamping the pattern of themes representing oppositeness subsides, leaving a new (consistent) vertical mapping in place, in which **abc** is seen as a successor-group going to the right and **xyz** as a predecessor-group going to the left. This way of looking at things makes sense with respect to the

wyz answer, since **c** and **x** are seen as corresponding. In this way, themes allow Metacat to effectively work backwards from a given answer to a high-level understanding of why the answer makes sense.

In conclusion, Metacat's themes can be viewed as a medium through which ideas made explicit at the "cognitive" level can actively influence and guide the course of processing at the "subcognitive" level. By strongly activating different patterns of themes in the Themespace, the program can explicitly focus on different high-level ideas as it works on understanding an analogy problem. Furthermore, once an answer has been understood, its associated themes represent a characterization of the key ideas underlying the answer, which can subsequently be used as the basis for comparing and contrasting the answer with other answers encountered previously.

ACKNOWLEDGEMENTS

This research is supported in part by Sun Microsystems Co. Academic Equipment Grant #EDUD-NAFO-960418 and by grants to the Center for Research on Concepts and Cognition from the College of Arts and Sciences of Indiana University.

REFERENCES

[Chi et al., 1989] Chi, M., Bassok, M., Lewis, M., Reimann, P., and Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13:145-182.

[Hofstadter and FARG, 1995] Hofstadter, D. R. and FARG (1995). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought.* Basic Books, New York.

[Mitchell, 1993] Mitchell, M. (1993). *Analogy-making as Perception.* MIT Press/ Bradford Books, Cambridge, MA.

[VanLehn et al., 1992] VanLehn, K., Jones, R., and Chi, M. (1992). A model of the self-explanation effect. *The Journal of the Learning Sciences*, 2(1):1-59.

MAPPING AND ACCESS IN ANALOGY-MAKING: INDEPENDENT OR INTERACTIVE? A SIMULATION EXPERIMENT WITH AMBR

Alexander A. Petrov[1], Boicho N. Kokinov[1,2]
[1] New Bulgarian University Department of Cognitive Science 21, Montevideo Str.
Sofia 1635, Bulgaria
apetrov@cogs.nbu.acad.bg
[2] Institute of Mathematics and Informatics, Bulgarian Academy of Science Bl. 8 Acad.
G. Bonchev Str.Sofia 1113, Bulgaria
kokinov@cogs.nbu.acad.bg

ABSTRACT

This paper contrasts two views about the relationship between the processes of access and mapping in analogy-making. According to the modular view, analog access and mapping are two separate 'phases' that run sequentially and relatively independently. The interactionist view assumes that they are interdependent subprocesses that run in parallel. The paper argues in favor of the second view and presents a simulation experiment demonstrating its advantages. The experiment is performed with the computational model AMBR and illustrates one particular way in which the subprocess of mapping can influence the subprocess of access.

INTRODUCTION

A crucial point in analogy-making is the retrieval of a base (or source) analog. Accessing an appropriate base from the vast pool of episodes stored in the long-term memory is not only a logical necessity (one cannot make analogies without a source) but apparently is the most difficult and capricious element of analogy-making. Starting with the classical experiments of Gick and Holyoak (1980) it has been repeatedly demonstrated that people have difficulties in spontaneously accessing a base analog, especially when its domain is very different from that of the target problem. In the aforementioned study only about 20% of the subjects were able to solve the so-called radiation problem even though an analogous problem (with solution) was presented shortly before the test phase. When provided by an explicit hint to use this source analog, however, 75% of the subjects achieved the solution. This great difference between the two experimental conditions was attributed to the difficulty of analog access.

On the other hand, we know a lot of stories about great scientists making discoveries by spontaneously using remote analogies. We have also personal experience in everyday usage of remote analogies. A recent study by Wharton, Holyoak, and Lange (1996) has demonstrated that about 35% of their subjects were successfully reminded about a remote analog story studied 7 days earlier when cued by the target story. (They have used a directed reminding task, not a problem solving task, however.)

Researchers of analogical access have become interested in the features of a remote analog that facilitate retrieval. Most data in the field (Holyoak and Koh, 1987, Ross 1989) suggest that analogical access is almost exclusively guided by superficial semantic similarities between

---

Keywords: Analogy-making, interactionist approach, access, mapping, simulation experiment, hybrid cognitive architecture.

base and target—similar objects and relations, similar themes, similar story lines, etc. In contrast, analogical mapping is dominated by the structural similarity between target and base, i.e. having common systems of relations (Gentner, 1983, 1989). This explains why remote analogs are much more difficult to access than to map—they lack the superficial similarities needed for access but do have the (quasi)isomorphic relational structure necessary for mapping.

This clear separation stimulated the researchers in the field to build separate models of mapping and retrieval and even to claim that they are different cognitive modules. Thus Gentner (1989) claims that 'the analogy processor (the mapping machine) is a well-defined separate cognitive module whose results interact with other processes, analogous to the way some natural language models have postulated semi-autonomous interacting subsystems for syntax, semantics, and pragmatics.' Although she explicitly mentions in a footnote that this should not be considered in the Fodorian sense as innate and impenetrable, the actual models built are quite impenetrable. This line of research has generated a number of quite successful models that explained the data and made some new predictions. Typically, a model of mapping is coupled with a (separate) model of retrieval. The best-known examples are SME + MAC/FAC (Falkenhainer, Forbus, and Gentner, 1986; Forbus, Gentner, and Law, 1995) and ACME + ARCS (Holyoak and Thagard, 1989; Thagard, Holyoak, Nelson, and Gochfeld, 1990).

However, the experimental work soon revealed that the pattern is not that clear and straightforward. It has been demonstrated that superficial similarities do play an important role in mapping as well. In particular cross-mapping is difficult (Ross, 1989). This led Holyoak and Thagard to include syntactic, semantic, and pragmatic constraints in their model of mapping ACME (Holyoak & Thagard, 1989) and to develop their multi-constraint theory (Holyoak & Thagard, 1995).

There are also some indications that structural similarity might play a role in access as well. Thus Ross (1989) demonstrated that in some cases (when the general story line is similar) structural similarity plays a positive role in retrieval, while in other cases (when the general story line is dissimilar) it does not play any role or can even worsen the results. The results of Wharton, Holyoak, and Lange (1996) also support indirectly the hypothesis that structural correspondences might affect the access. This was reflected in the models being proposed. Both MAC/FAC and ARCS included a submodule of partial mapping in the module of retrieval, thus considering structural similarities at an early stage.

To sum up, the initial separation between retrieval and mapping was founded on their different psychological characteristics—semantic factors govern the retrieval, structural factors govern the mapping. Subsequent more precise experiments, however, cast doubt on this clear separation. These complications were accommodated by making patches to the original models. Finally, it was acknowledged that all kinds of constraints affected all phases of analogy-making, although to different extent (Holyoak & Thagard, 1995).

The experimental data themselves became more and more complex and controversial. These controversies can be explained in terms of more and more sophisticated classifications of the types of similarities involved in access and mapping (Ross, 1989; Ross & Kilbane, 1997). We argue, however, that these problems are resolved more parsimoniously by adopting a principally different view of analogy-making.

This resembles an episode of the history of astronomy. The geocentric system of Ptolemy started as a straightforward theory that described the observable movement of both stars and planets remarkably well. As accuracy of measurement increased, however, discrepancies between theory and data crept in every now and then. It became routine for astronomers to deal with such 'anomalies' by adding more and more epicycles. But as time went on, it became evident that astronomy's complexity was increasing far more rapidly than its accuracy and that a discrepancy corrected in one place was likely to show up in another (Kuhn, 1970).

Back to the domain of analogy-making, most classical models assume sequential processing: *first* the retrieval process finds the base for analogy and *then* the mapping process builds the correspondences between the target and the retrieved base (Figure 1). Thus MAC/FAC+SME and ARCS+ACME are linear models separating retrieval and mapping in time and space. This view underlies most of the experimental work in the field as well. Researchers often contrast hint versus non-hint conditions in problem solving supposing that in the first case only mapping takes place, while in the second retrieval and mapping are running one after the other. However, as Ross (1989) has noted, even when explicitly hinted to use a certain analog subjects still must access the details of its representation. Another common experimental technique uses a memory task (typically recall) for studying access with the assumption that the same processes take place during analogical problem solving.

The limitations of both the models and experimental methods can be overcome by giving up the linearity assumption. This might look strange at first glance—how can you map the source analog onto the base if you have not even accessed it?! If, however, one reconsiders one more assumption—that there are centralized representations of situations/problems in human memory—then it becomes clear that whenever we have partial retrieval of the base (having recalled a few details) we can start looking for corresponding elements in the target. This allows us to conceptualize access and mapping as parallel processes that can interact (Figure 2). In this paradigm, access and mapping refer not to phases or other behavioral steps, but rather to separate mechanisms that both play a role in selecting and activating a base and in finding the correspondences between base and target.

The current paper explores the implications of the parallel and interactive view on access and mapping by running simulation experiments with an integrated model of human (analogical) reasoning called AMBR (Kokinov, 1988, 1994c, Petrov, 1997). These experiments provide a detailed example of how these two processes can interact and thus open space for new theoretical speculations as well as for new experimental paradigms. AMBRÓs predictions about the development of the process over time call for appropriate experimental methods capturing the dynamics of human analogy-making—RT studies, think-aloud protocols, etc. Some of the controversies around the role of superficial and structural similarities in access and mapping 'phases' can now be expressed in terms of the interactions between the two mechanisms.

A very important contribution of the simulation is that it demonstrates how the supposedly later 'phase' of mapping can influence the supposedly earlier 'phase' of access. A detailed example shows how the access process develops over time and how it is influenced by the concurrent mapping process. This is contrasted with the case of isolated access. Different results are obtained in the two cases. These results correspond to the data of Ross and Sofka (unpublished) which main conclusions are summarized in (Ross, 1989) as follows: '... other work (Ross & Sofka, 1986) suggests the possibility that the retrieval may be greatly affected by the use.

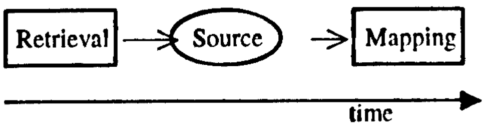


Figure 1. Dominating sequential models of analogy-making.

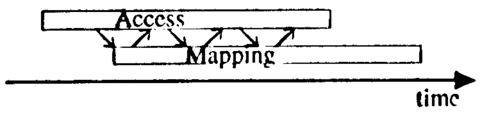


Figure 2. Parallel and interactive models of analogy-making.

In particular, we found that subjects, whose task was to recall the details of an earlier example that the current test problem reminded them of, used the test problem not only as an initial reminder but throughout the recall. For instance, the test problem was used to probe for similar objects, and relations and to prompt recall of particular numbers from the earlier example. The retrieval of the earlier example appeared to be interleaved with its use because subjects were setting up correspondences between the earlier example and the test problem during the retrieval.' The simulation data presented in the current paper (obtained absolutely independently and based only on the theoretical assumptions of DUAL and AMBR) exhibit exactly the same pattern of interaction.

We must admit that even in a highly parallel and interactive model such as AMBR the effects of interactions are not predominating. In the majority of cases the independent work of the access mechanism might well yield the same results as the interaction between mapping and access described above. That is why the classical linear models of analogy have been successful and have contributed a lot to our understanding of human analogy-making. However, exactly the few exceptional cases that do provide different results in a parallel model are the more interesting and those who make the interpretation of the experimental data look controversial if analyzed in the spirit of the sequential models.

There are a few other models that advocate a parallel, overlapping, and interactive view on analogy—Copycat (Mitchell, 1993, Hofstadter, 1995), Tabletop (French, 1995, Hofstadter, 1995), and LISA (Hummel and Holyoak, 1997). However, Copycat and Tabletop do not model retrieval at all—they model the parallel work and interaction between perception/representation building and mapping. LISA also integrates access and mapping and performs them in parallel. Thus the mapping mechanism (connectionist learning in this case) influences the access. As a result, LISA could in principle demonstrate effects similar to those reported here.

BRIEF DESCRIPTION OF THE ARCHITECTURE DUAL AND THE MODEL AMBR

The basis for the simulation experiment discussed in this paper is a model called AMBR (Associative Memory-Based Reasoning). It is built on the cognitive architecture DUAL. Space limitations allow only an extremely sketchy description of DUAL and AMBR here. The interested reader is referred to earlier publications (Kokinov, 1988, 1994a,b,c; Petrov, 1997).

DUAL is a multi-agent cognitive architecture that supports dynamic emergent computation (Kokinov,Nikolov, and Petrov, 1996). All knowledge representation and information processing in the architecture is carried out by small entities called *DUAL agents*. Each DUAL-based system consists of a large number of them. There is no central executive in the architecture that controls its global operation. Instead, each individual agent is relatively simple and has access only to local information, interacting with a few neighboring agents. The overall behavior of the system emerges out of the collective activity of the whole population. This 'society of mind' (Minsky, 1986) provides a substrate for concurrent processing, interaction, and emergent computation.

Each DUAL agent is a hybrid entity that has symbolic and connectionist aspects (Kokinov 1994a,b,c). On the symbolic side, each agent 'stands for' something and is able to perform certain simple manipulations on symbols. On the connectionist side, it sends/receives activation to and from its immediate neighbors. Thus, we may adopt an alternative terminology and speak of *nodes* and *links* instead of *agents* and *interactions*. The population of agents may be conceptualized as a network of nodes.

The long-term memory of a DUAL-based system consists of the network of all agents in that system. The size of this network can be very large. Only a small fraction of it, however, may be active at any particular moment. The active subset of the long-term memory together with some temporary agents constitutes the *working memory (WM)* of the architecture. The mechanism of spreading activation plays a key role for controlling the size and the contents of the WM. There is a threshold that sets the minimal level of activation that must be obtained by an agent to enter the WM. There is also a spontaneous decay factor that pushes the activation levels back to zero. As the pattern of activation changes over time, some agents from the working memory fall back to dormancy, others are activated, etc. Only active agents may perform symbolic computation. Moreover, the speed of this computation depends on the level of activation of the respective agent. This makes the computation in DUAL dynamic and context-sensitive (Kokinov et al., 1996; Kokinov, 1994a,b,c). One particular consequence of this dynamic emergent nature of the architecture is that, although all micro-level processing is strictly deterministic, the macroscopic behavior of a DUAL system can be described only probabilistically.

The AMBR model takes advantage of these architectural features to account for some phenomena of human reasoning and in particular reasoning by analogy (Kokinov, 1988, 1994c). Again, due to space limitations we will consider only a small fraction of model's mechanisms.

Analog access in AMBR is done by means of spreading activation by the connectionist aspects of the DUAL agents. In particular, only few of the many episodes stored in the long-term memory are active during a run and only they are accessible for processing. The episodes or 'situations' have decentralized representations—it is not a single agent but a whole *coalition* that represents the elements of a situation and the relationships among them. Therefore, it is possible that an episode is only partially accessed because only some of the agents have entered the WM.

The process of analogical mapping is done in AMBR by a combination of three mechanisms—marker passing, constraint satisfaction, and structure correspondence (Kokinov, 1994c; Petrov, 1997). The main idea is to build a *constraint satisfaction network (CSN)* to determine the mapping between two situations. This network consists of *hypothesis agents* representing tentative correspondences between two elements. Consistent hypotheses support, and incompatible ones inhibit each other.

This is similar to other models of analogy-making and notably ACME (Holyoak and Thagard, 1989). AMBR differs from the latter model, however, in several ways: (*i*) the CSN is constructed dynamically, (*ii*) only hypotheses that have some justification are created, (*iii*) the CSN is incorporated into the bigger working memory network, and (*iv*) there is no separate relaxation phase so there is a partial mapping at each moment.

The implication of these four points is that, unlike ACME and most other analogy models, the processes of access and mapping run in parallel and influence each other in AMBR. In other words, the model departs from the classical 'pipeline' paradigm and aims at a more interactive account of analogy making.

The influence between the two subprocesses in AMBR goes in both directions. The present paper concentrates on the 'backward' direction—from mapping to access. The next section describes a simulation experiment that sheds light on this kind of influence.

SIMULATION EXPERIMENT METHOD

We performed a simulation experiment to contrast the two ways of combining access and mapping—parallel vs. serial. The experiment also tested whether the AMBR model was capable to access a source analog out of a pool of episodes, and to map it onto a target situation.

Design

The experiment consisted of two conditions. Both conditions involved running the model on a target problem. In the 'parallel condition', AMBR operated in its normal manner with the mechanisms for access and mapping working in parallel. In the 'serial condition', the program was artificially forced to work serially—first to access and only then to map. The target problem and the content of the long-term memory were identical in all runs. The topics of interest fell into two categories—the final mapping constructed by the program and the dynamics of the underlying computation. The latter was monitored by recording a set of variables describing the internal state of the system at regular time intervals throughout each run.

Materials

The domain used in the experiment deals with simple tasks in a kitchen. The long-term memory of the model contains semantic and episodic knowledge about this domain. It has been coded by hand according to the representation scheme used in Dual and Ambr (Kokinov, 1994c; Petrov, 1997). The total size of the knowledge base is about 500 agents (300 'semantic' + 200 'episodic'). It states, for example, that water, milk, and tea are all liquids, that bottles are made of glass, and the relation 'on' is a special case of 'in-touch-with'. The LTM also stores the representations of eight situations related to heating and cooling liquids. Two of these eight situations are most important for the experiment and are described below together with the target problem.

As evident from Figures 3, 4, and 5, both situations **A** and **B** may be considered similar to the target problem. There are some differences, however. Situation **B** involves the same objects and relations as the target but the structure of the two are different. In contrast, situation **A** involves different objects but its system of relations is completely isomorphic to that of the target. According to Gentner (1989), the pair **A-T** may be classified as analogy while **B-T** as mere appearance. Thus it was expected that situation **B** would be easier to retrieve from the total pool of episodes stored in LTM. On the other hand, **A** would be more problematic to retrieve but once accessed it would support better mapping.

Situation A: *There is a cup and some water in it. The cup is on a saucer and is made of china. There is an immersion heater in the water. The immersion heater is hot. The goal is that the water is hot.*

The outcome is that the water is hot. This is caused by the hot immersion heater in it.

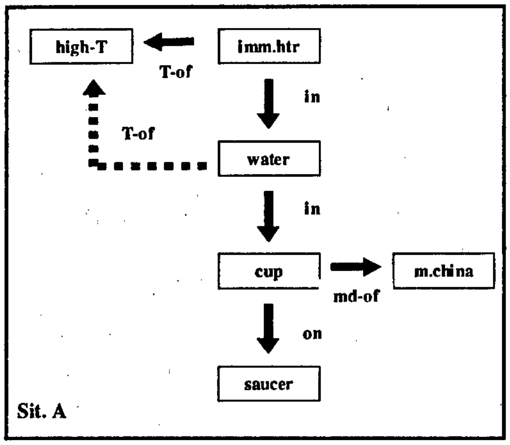


Figure 3. Schematized representation of situation A. Objects are shown as boxes and relations as arrows. Dashed arrows stand for relations in the 'outcome'. The actual Ambr representation is more complex—it consists of 19 agents and explicates the causal structure (not shown in the figure). See text for details.

Situation B: *There is a glass and an ice cube on it. The glass is made of [material] glass. The glass is in a fridge. The fridge is cold The goal is that the ice cube is cold.*

The outcome is that the ice cube is cold. The fact that it is on the glass and the glass is in the fridge entails that the ice cube itself is in the fridge. In turn, this causes the ice cube to be cold, as the fridge is cold.

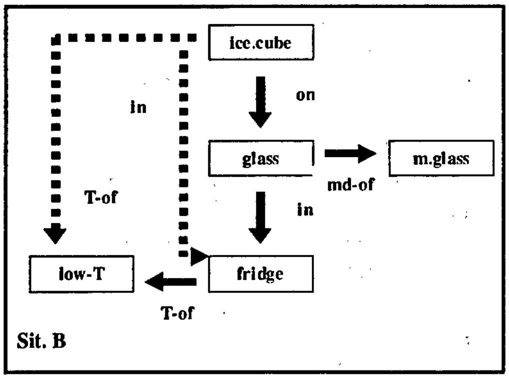


Figure 4. Schematized representation of situation B. Dashed arrows stand for relations in the 'outcome'. The actual Ambr representation is more complex—it consists of 21 agents and explicates the causal structure (not shown in the figure). See text for details.

Target problem (situation T): *There is a glass and some coke in it. The glass is on a table and is made of [material] glass. There is an ice cube in the coke. The ice cube is cold. The goal, if any, is not represented explicitly.*

What is the outcome of this state of affairs?

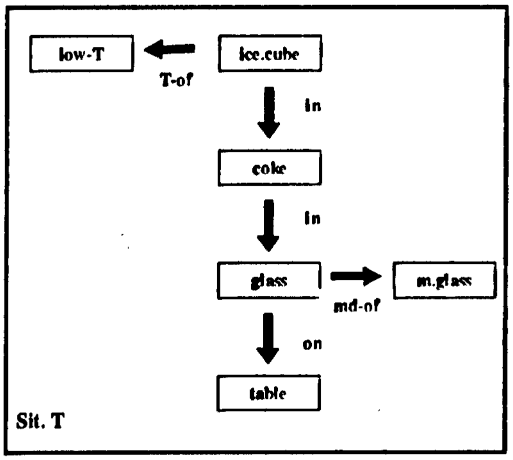


Figure 5. Schematized representation of the target situation. The actual Ambr *representation is more complex and consists of 15 agents. See text for details.*

Procedure

The Common Lisp implementation of the Ambr model was run two times on the target problem. The two runs carried out the 'parallel' and the 'serial' conditions of the experiment, respectively. The contents of the long-term memory and the parameters of the model were identical in the two conditions.

Recall that situations have decentralized representations in Ambr. The target problem was represented by a coalition of 15 agents standing for the ice-cube, the glass, two instances of the relation 'in' and so on. 12 of these agents were attached to the special nodes that serve as activation sources in the model. The attachment was the same in the two experimental conditions.

In the parallel condition, the model was allowed to run according to its specification. That is, all Ambr mechanisms ran in parallel, interacting with one another. The program iterated until the system reached a resting state. A number of variables were recorded at regular intervals throughout the run. Out of these many variables, the so-called *retrieval index* is of special interest. It is computed as the average activation level of the agents involved in each situation.

In short, at the end of the run we had the final mapping constructed by the program as well as a log file of the retrieval indices of all eight situations from the LTM.

In the serial condition, the target problem was attached to the activation source in the same way and the same data were collected. However, the operation of the program was forcefully modified to separate the processes of access and mapping. To that end, the run was divided in two steps.

During step one, all mapping mechanisms in Ambr were manually switched off. Thus, spreading activation was the only mechanism that remained operational. It was allowed to work until the pattern of activation reached asymptote. The situation with the highest retrieval index was then identified. If we hypothesize a 'retrieval module', this is the situation that it would access from LTM.

After the source analog was picked up in this way, the experiment proceeded with step two. The mapping mechanism was switched back on again but it was allowed to work only on the source situation retrieved at step one. This situation was mapped to the target. Thus, at the end of the second run we had the final mapping constructed at step two, as well as two logs of the retrieval indices.

RESULTS AND DISCUSSION

In both experimental conditions the model settled in less than 150 time units and produced consistent mappings. By 'consistent' we mean that each element of the target problem was unambiguously mapped to an element from LTM and that all these corresponding elements belonged to one and the same base situation. Stated differently, the mappings were one-to-one and there were no blends between situations.

In the parallel condition, the target problem was mapped to situation **A**, yielding the correspondences *in–in, water–coke, imm.heater–ice.cube, T.of–T.of, high.T–low.T,* *made.of–made.of*, etc. Four elements from the source situation remained unmapped and in particular the agent representing that the water is hot. This proposition is a good candidate for inference by analogy. *Mutatis mutandis*, it could bring the conclusion that the coke is cold. (In the current version of AMBR2 the mechanisms for analogical transfer are not implemented yet.)

In the serial condition, situation **B** won the retrieval stage. This is explained by the high semantic similarity between its elements and those of the target—both deal with ice cubes in glasses, cold temperatures, etc. The asymptotic level of the retrieval index for **B** was about four times greater than that of any other situation. In particular, situation **A** ended up with only 5 out of 19 agents passing the working memory threshold.

According to the experimental procedure, situation **B** was then mapped to the target during the second stage of the run. The correspondences that emerged during the latter stage are shown in Table 1. The semantic similarity constraint has dominated this run. This is not surprising given the high degree of superficial similarity between the two situations. There is, however, a serious flaw in the set of correspondences. The proposition 'T.of (ice.cube, low-T)', which belongs to the *initial* state of the target, is mapped to the proposition 'T.of (ice.cube, low-T)', which is a *consequence* in the source. Therefore, the whole analogy between the target problem and the situation **B** could hardly generate any useful inference.

Situation B	Target situation
ice.cube	ice.cube
fridge	coke
glass	glass
in (ice.cube, fridge)	in (ice.cube, coke)
in (glass, fridge)	in (coke, glass)
on (ice.cube, glass)	on (glass, saucer)
T.of (fridge, low-T)	<unmapped>
T.of (ice.cube, low-T)	T.of (ice.cube, low-T)
low-T	low-T
made.of (glass, m.glass)	made.of (glass, m.glass)
m.glass	m.glass
initstate1	initstate
initstate2	<unmapped>
interstate	table
endstate	endstate
goalstate	<unmapped>
follows (initstate1, endst.)	follows (initstate, endst.)
to.reach (initstate1, goalst)	<unmapped>
cause (initstate2, in(i.c,fr))	<unmapped>
cause (interstate, T.of(i.c))	<unmapped>

Table 1. Correspondences constructed by the model in the serial condition.

To summarize, when the mechanisms for access and mapping worked together, the model constructed an analogy that can potentially solve the problem. On the other hand, when the two mechanisms were separated, the retrieval stage favored a superficially similar but inappropriate base.

The presentation so far concentrated on the final set of correspondences produced by the model. We now turn to the dynamics of the computation as revealed by the time course of the retrieval indices. Figure 6 plots the retrieval indices for several LTM episodes during the first run of the program (i.e. when access and mapping worked in parallel). Figure 7 concentrates on the early stage of the first run and compares it with the second run (i.e. when only the access mechanism was allowed to work). Note that the two plots are in different scales.

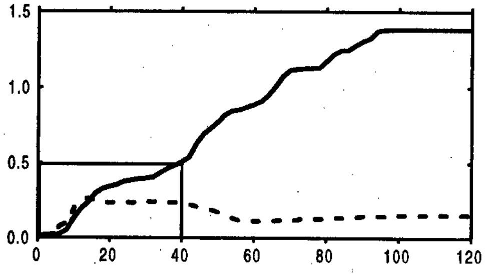


Figure 6. Plot of retrieval indices versus time for the parallel condition. Situation A is in solid line, B in dashed. The 'south-west' corner of the plot is reproduced in Figure 7 with threefold magnification.

These plots tell the following story: At the beginning of the parallel run, several situations were probed tentatively by bringing a few elements from each into the working memory. Of this lot, **B** looked more promising than any of its rivals as it had so many objects and relations in common with the target. Therefore, about half of the agents belonging to situation **B** entered the working memory and began trying to establish correspondences between themselves and the target agents. The active members of the rival situations were doing the same thing, although with lower intensity. At about 15 time units since the beginning of the simulation, however, situation **A** (with the immersion heater) rapidly gained strength and eventually overtook the original leader. At time 40, it had already emerged as winner and gradually strengthened its dominance.

The final victory of situation **A**, despite its lower semantic similarity compared to situation **B**, is due to the interaction between the mechanisms of access and mapping in AMBR. More precisely, in this particular case it is the mapping that radically changes the course of access. To illustrate the importance of this influence, Figure 7 contrasts the retrieval indices with and without mapping.

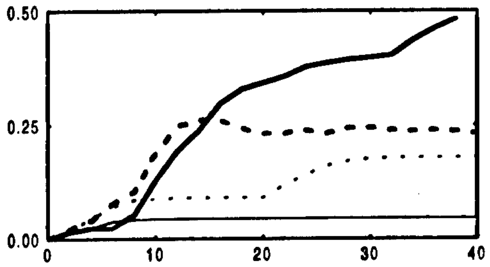


Figure 7. Retrieval indices for situations A and B with and without mapping influence on access. The thick lines correspond to the parallel condition and replicate (with threefold magnification) the lines from the 'southwest' corner of Fig. 6. The thin lines show 'pure' retrieval indices. See text for details.

The thin lines in Figure 7 show the retrieval indices for the two situations when mapping mechanisms are suppressed. Thus, they indicate the 'pure' retrieval index of each situation—the value that is due to the access mechanism alone. The index for situation **B** is much higher than that of **A** and, therefore, **B** was used as source when the mapping was allowed to run only after the access had finished.

The step-like increases of the plots indicate moments in which an agent (or usually a tight sub-coalition of two or three agents) passes the working memory threshold. This happens, for example, with situation **B** between time 20 and 30 of the serial condition (the thin dashed line in Figure 7). Thus, accessing a source episode in AMBR is not an all-or-nothing affair. Instead, situations enter the working memory agent by agent and this process extends far after the beginning of the mapping. In this way, not only can the access influence the mapping but also the other way around.

In the interactive condition the mapping mechanism boosted the retrieval index via what we call a 'bootstrap cascade'. This cascade operates in AMBR in the following way. First, the access mechanism brings two or three agents of a given situation into the working memory. If the mapping mechanism then detects that these few agents can be plausibly mapped to some target elements, it constructs new correspondence nodes and links in the AMBR network. This creates new paths for the highly active target elements to activate their mates. The latter in turn can then activate their 'coalition partners', thus bringing a few more agents into the working memory and so on.

The bootstrap cascade is possible in AMBR due to two important characteristics of this model. First, situations have decentralized representations which may be accessed piece by piece. Second, AMBR is based on a parallel cognitive architecture which provides for concurrent operation of numerous interacting processes. Taken together, these two factors enable seamless integration of the subprocesses of access and mapping in analogy-making.

CONCLUSION

The simulation experiment reported in this paper provides a clear example of mapping influence on analog access and of the advantages of the parallel interactionist view on analogy-making. Furthermore, the computational model AMBR provides a theoretical framework for explaining the controversies in the psychological data on access and reminding. It is possible to explore in which cases the interaction between access and mapping produces results different from a sequential and independent processing. It provides also a framework for generating more precise hypotheses and new experimental designs for their testing. Thus, for example, the detailed logs of the running model might be used for comparison with protocols of think-aloud experiments.

Analogy-making has certainly no clear cut boundaries. Most literature has concentrated on explicit analogies, i.e. consciously retrieving an analog and noticing the analogy. However, there are other cases which might be called implicit or partial analogies, e.g. subconsciously accessing part of a previously solved problem and mapping it to part of the target description without consciously noticing the analogy. The decentralized representations of situations in AMBR make it possible to model the process of partial access, access with distortions, blending (Turner & Fauconnier, 1995), and interference. A previously solved problem can influence the course of problem solving in an even more subtle way by priming some concepts or situations which then trigger a particular solution (Kokinov, 1990, Schunn and Dunbar, 1996). The AMBR model can be used to analyze such cases. It has already been successfully applied for predicting priming and context effects (Kokinov, 1994c).

Priming effects are an example of the influence of access on mapping which is the opposite direction of the one discussed in the current paper. Order effects are another kind of effect that goes in 'forward' direction. Such effects may be due to non-simultaneous perception of the elements of the target problem (Keane, Ledgeway, & Duff, 1994) and/or non-simultaneous retrieval of relevant pieces of information from LTM. Thus the mutual influence between analog access and mapping offers many opportunities for investigation.

REFERENCES

Falkenhainer, B., Forbus, K., and Gentner, D. (1986). The structure-mapping engine. *Proceedings of the Fifth Annual Conference on Artificial Intelligence*. Los Altos, CA: Morgan Kaufman.

Forbus K., Gentner D., and Law, K (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141-205.

French, R. (1995). *The subtlety of sameness: A theory and computer model of analogy-making*. Cambridge, MA: MIT Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou and A. Ortony (Eds.), *Simiarity and analogical reasoning*. New York, NY: Cambridge University Press.

Gick, M.L. and Holyoak, K.J. (1980). Analogical problem solving. *Cognitive Psychology 12* (80), 306-356.

Hofstadter, D. and the Fluid Analogies Research Group (1995). *Fluid concepts and creative analogies: Comuter models of the fundamental mechanisms of thought.* New York: Basic Books.

Holyoak K. and Koh K. (1987). Surface and structural similarity in analogical transfer. *Memory and Cognition, 15* (4), 332-340.

Holyoak K. and Thagard P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295-355.

Holyoak, K. and Thagard, P. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

Hummel, J. and Holyoak, K. (1997). Distributed representation of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427-466.

Keane, M., Ledgeway, K., and Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science, 18*, 387-438.

Kokinov, B. (1988). Associative memory-based reasoning: How to represent and retrieve cases. In T. O'Shea and V. Sgurev (Eds.), *Artificial intelligence III: Methodology, systems, applications*. Amsterdam: Elsevier.

Kokinov, B. (1990). Associative memory-based reasoning: Some experimental results. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kokinov, B. (1994a). The context-sensitive cognitive architecture DUAL. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale,NJ: Lawrence Erlbaum Associates.

Kokinov, B. (1994b). The DUAL cognitive architecture: A hybrid multi-agent approach. *Proceedings of the Eleventh European Conference of Artificial Intelligence*. London: John Wiley & Sons, Ltd.

Kokinov, B. (1994c). A hybrid model of reasoning by Analogy. In K. Holyoak and J. Barnden (Eds.), *Advances in Connectionist and Neural Computation Theory. Vol. 2: Analogical Connections*. Norwood, NJ: Ablex Publishing Corp.

Kokinov,B., Nikolov,V., and Petrov,A. (1996). Dynamics of emergent computation in DUAL. In A. Ramsay (Ed.), *Artificial Intelligence: Methodology, Systems, Applications*. Amsterdam: IOS Press.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions* (second ed., enlarged). Chicago: The University of Chicago Press. (First edition published 1962.)

Minsky, M. (1986). *The society of mind*. New York: Simon and Schuster.

Mitchell, M. (1993). *Analogy-making as perception: A computer model*. Cambridge, MA: MIT Press.

Petrov, A. (1997). *Extensions of DUAL and AMBR*. M.Sc. Thesis. New Bulgarian University, Cognitive Science Department.

Ross, B. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 456-468.

Ross, B. and Kilbane, M. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23* (2), 427-440.

Ross, B. and Sofka, M. (1986). [Remindings: Noticing, remembering, and using specific knowledge of earlier problems]. Unpublished manuscript.

Schunn, C. and Dunbar, K. (1996). Priming, analogy, and awareness in complex reasoning. *Memory and Cognition, 24*, 271-284.

Thagard, P., Holyoak, K., Nelson, G., and Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence, 46*, 259-310.

Turner, M. and Fauconnier, G. (1995). Conceptual integration and formal expression. *Metaphor and Symbolic Activity, 10* (3), 183-204.

Wharton, C., Holyoak, K., and Lange, T. (1996). Remote analogical reminding. *Memory and Cognition, 24* (5), 629-643.

PRINCIPLE DIFFERENCES IN STRUCTURE-MAPPING: *THE CONTRIBUTION OF NON-CLASSICAL MODELS*

Tony Veale
School of Computer Applications.
Dublin City University,
Dublin 9, Ireland.

Mark T. Keane
Dept. of Computer Science,
University College Dublin,
Dublin 2, Ireland.

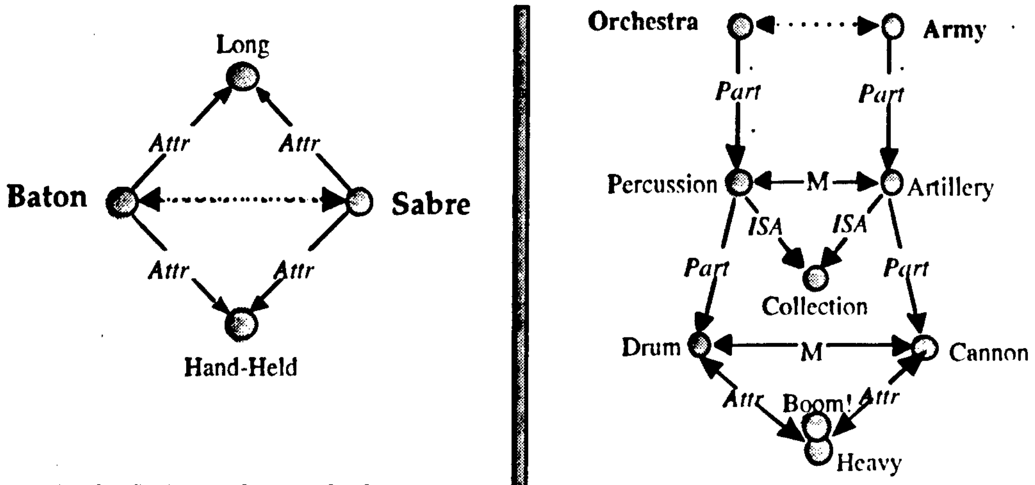
ABSTRACT

Recent research in metaphor and analogy, as variously embodied in such systems as Sapper, LISA, Copycat and TableTop, speak to the importance of three principles of cross-domain mapping that have received limited attention in, what might be termed, the *classical analogy* literature. These principles are that: (i) high-level analogies arise out of nascent, lower-level analogies automatically recognized by memory processes; (ii) analogy is memory-situated inasmuch as it occurs *in situ* within the vast interconnected tapestry of long-term semantic memory, and may potentially draw upon any knowledge fragment; and (iii), this memory-situatedness frequently makes analogy necessarily dependent on some form of attributive grounding to secure its analogical interpretations. In this paper we discuss various arguments, pro and con, for the computational and cognitive reality of these principles.

INTRODUCTION

Over the last few years, we have been examining the computational capabilities of models of analogy (see Veale & Keane, 1993, 1994, 1997; Veale *et al.*, 1996). Some models of analogy, like the original version of the Structure-Mapping Engine (SME; Falkenhainer, Forbus & Gentner, 1989), have been concerned with producing optimal solutions to the computational problems of structure mapping, although more recently, many models have adopted a more heuristic approach to improve performance at the expense of optimality; models like the Incremental Analogy Machine (IAM; Keane & Brayshaw, 1988; Keane *et al.*, 1994), the Analogical Constraint Mapping Engine (ACME; Holyoak & Thagard, 1989), Greedy-SME (see Forbus & Oblinger, 1990) and Incremental-SME (see Forbus, Ferguson & Gentner, 1994). These *classical structure-mapping models* have also been predominantly concerned with modelling the details of a corpus of psychological studies on analogy.

In contrast, there is a different *non-classical* tradition that has concentrated on capturing key properties of analogising, with less reference to the mainstream psychological literature (e.g., the Copycat system of Hoftstadter *et al.* 1995; the TableTop system of Hofstadter & French, 1995; and the AMBR system of Kokinov, 1994). Recently, there has been something of a confluence of these two traditions as models have emerged that exhibit many of the parallel processing properties of non-classical approach with the computational and empirical constraints of classical models; models like Sapper (see Veale & Keane, 1993, 1994, 1997; Veale *et al.*, 1996) and LISA (see Hummel and Holyoak, 1997). While these models are clearly differ-



(i) The Triangulation Rule

(ii) The Squaring Rule

Figure 1. The Triangulation Rule (i) and the Squaring Rule (ii) augment semantic memory with additional bridges (denoted *M*), indicating potential future mappings.

ent to classical models, it is not immediately obvious whether they are just algorithmic variations on the same computational-level theme, or whether they constitute a significant departure regarding the *principles of analogy*. In this paper, using Sapper as a focus, we argue that there are at least three principles on which Sapper differs from wholly classical models. We also argue from a computational perspective that Sapper offers several performance efficiencies over optimal and sub-optimal classical models.

PRINCIPAL DIFFERENCES

Sapper accepts most of the computational-level assertions made about structure mapping, such as the importance of isomorphism, structural consistency and systematicity (see Keane *et al.*, 1994, for a computational-level account). A ongoing discussion with several researchers in the field has helped to define its differences in-principle from classical models (c.f. Ferguson, Forbus & Gentner, 1997; Thagard, 1997). In summary, they are that:

- *Analogies are forever nascent in human memory*: that human memory is continually preparing for future analogies by establishing potential mappings between domains of knowledge.
- *Mapping is memory-situated*: that mapping occurs within a richly elaborated, tangle of conceptual knowledge in long-term memory.
- *Attributes are important to mapping*: that attribute/category information plays a crucial role in securing both the relevance *and* tractability of an analogical mapping.

At present, the psychological literature is silent on many of these points. In this paper, we address these issues by outlining each of the principles in more detail and evaluating the computational and psychological evidence of relevance to them.

NASCENT ANALOGIES

The picture Sapper creates of the analogy process is quite different from the goal-driven, *just-in-time* construction of analogies associated with the classical models. In the classical tradition, all analogising occurs when current processing demands it, a proposal that is most obvious in the centrality given to pragmatic constraints (see Holyoak & Thagard, 1989; Keane 1985; Forbus & Oblinger, 1990). In these models, mappings are constructed when the system goes into "analogy mode" and are not prepared in advance of an analogy-making session. In contrast, Sapper models analogy-making as a constant background activity where potential mappings are continually and pro-actively prepared in memory, to be exploited when particular processing goals demand them to be used. Analogies are thus forever nascent in Sapper's long-term memory.

Sapper forms analogies using spreading-activation within a semantic network model of long-term memory, by exploiting *conceptual bridges* that have been established between concepts in this network. These bridges record potential mappings between concepts and are automatically added by Sapper to its semantic network when the structural neighbourhoods of two concepts share some local regularity of structure. Such bridges are highly tentative when initially formed, and thus remain dormant inasmuch as they are not used by "normal" spreading activation in the network. But dormant bridges can be awakened, and subsequently used for spreading activation, when some proposed analogical correspondence between the concepts is made by the cognitive agent.

The regularities of structure which Sapper exploits to recognize new *bridge-sites* in long-term memory are captured in two rules that are graphically illustrated in Figure 1: the triangulation and squaring rules. The *triangulation rule* asserts that:

> **If memory already contains two linkages L_{ij} and L_{kj} of semantic type L forming two sides of a triangle between the concept nodes C_i, C_j and C_k, then complete the triangle and augment memory with a new bridge linkage B_{ik}.**

For example, in Figure 1(i), when concepts *BATON* and *SABRE* have the shared predicates LONG and HANDHELD the triangulation rule will add a bridge between them, which may subsequently be exploited by an analogy. In predicate calculus notation, this could be interpreted as asserting that when two concepts partake in two or more instances of predications which are otherwise identical, they become candidates for an analogical mapping, e.g., that *long(BATON) & handheld(BATON)* and *long(SABRE) & handheld(SABRE)* suggest that *BATON* and *SABRE* are candidates for an entity mapping in a later analogy. Memory is thus seen by Sapper as pro-actively exploiting perceptual similarities to pave the way for future structural analogies and metaphors; much like Hofstadter & French (1995) then, Sapper views analogy and metaphor as outcrops of low-level perception.

The structural integrity of these analogical outcrops is enforced by the *squaring rule,* which works at a higher level over collections of bridges between concepts:

> **If B_{ik} is a conceptual bridge, and if there already exists the linkages L_{mi} and L_{nk} of the predicate type L, forming three sides of a square between the concept nodes C_i, C_k, C_m and C_n, then complete the square and augment long-term memory with a new bridge linkage B_{mn}.**

For example, in Figure 1(ii) the bridges established using triangulation between *PERCUSSION -> ARTILLERY* and *DRUM -> CANNON*, support the formation of an additional bridge between *ORCHESTRA* and *ARMY* using the squaring rule. The intuition here is that correspondences based on low-level semantic features can support yet higher-level correspondences (see Hofstadter *et al.* 1995; Hummel & Holyoak, 1997).

The proposal that analogies are forever nascent in human memory may seem computationally implausible because it suggests a proliferation of conceptual bridges that would quickly overwhelm our memories with irrelevant conceptual structure. In practice, this does not seem to be the case. In performance experiments, we have shown that as a knowledge-base grows so too does the number of bridges, but in a polynomially modest fashion (see Veale *et al.* 1996).

Indeed, the notion of a conceptual bridge is a compelling one that seems to have emerged independently from multiple researchers in the field (e.g., Veale & Keane, 1993; Eskridge, 1994; Hofstadter *et al.*, 1995). From a psychological perspective, some have argued that forming potential mappings in advance of an analogy is implausible (e.g., see Ferguson *et al.*, 1997). While we know of no evidence that directly supports or denies the bridging stance, it does gel with certain broad phenomena. The inherent flexibility and speed of people's analogical mapping, even within relatively large domains, suggests that some pre-compiled correspondences are used, otherwise the mapping problem approaches intractability; this is especially so when slippage and re-representation in these domains is also implicated. Similarly, Hofstadter and his team's characterisation of people's alacrity in performing conceptual slippage between different entities is more consistent with this account than classical models would be.

MAPPING IS MEMORY-SITUATED

Sapper sees the mapping process as being essentially *memory-situated*, that is, that the generation of mapping-rich interpretations can only be carried out within a long-term memory of richly interconnected concepts. In character, this is quite different to classical models which see analogues as delineated bundles of knowledge, segregated parcels of predications that are retrieved from memory and mapped in "another place" (usually a temporary working memory). In some cases, this knowledge-bundling seems more plausible than in others. For instance, it makes some sense in the encoding of episodic event sequences (typically, used in bench-marking analogy models), although even in these cases many of the properties of object-centred concepts (i.e., those typically expressed at a linguistic level via nouns rather than verbs) seem to be unnaturally suppressed. This bundling makes less sense in other cases, as in the profession domains used in Sapper where objects (such as GENERAL, SURGEON, SCALPEL, ARMY, etc.) are the focal points of the representation, and relations are hung between them. In turn, this has led to the objection that Sapper's test domains inappropriately include "the whole of semantic memory" in the domain representation (c.f. Thagard, 1997). We would argue that this is entirely the point; natural analogy is performed within large, elaborated domains involving many predicates with few clear boundaries on relevance. Since clever analogies and metaphors surprise and delight us by the unexpected ways in which they relate the dissimilar, the mapping device is frequently itself the relevance mechanism. Let's consider then how Sapper forms analogies in a memory-situated fashion.

Sapper performs analogical mapping by spreading activation through its semantic memory, pin-pointing cross-domain bridges that might potentially contribute to a final interpretation (see Appendix A for the algorithm). The algorithm first performs a bi-directional breadth-first search from the root nodes of the source (S) and target (T) domains in memory, to seek out all relevant bridges that might potentially connect both domains and thus finds an intermediate set of candidate matches (or *pmaps*, in SME parlance). To avoid a combinatorial explosion, this search is limited to a fixed horizon H of relational links (usually H = 6) while employing the same predicate identicality constraint as SME for determining structural isomorphism. Then, the richest pmap (i.e., the pmap containing the largest number of cross-domain mappings) is chosen as a seed to anchor the overall interpretation, while other pmaps are folded into this seed if they are consistent with the evolving interpretation, in descending order of the richness of those pmaps (in a manner that corresponds closely to Greedy-SME)[1]. The use of *memory-situatedness* in combination with the other features of Sapper delivers effective performance on mapping these analogies.

Tests of Sapper relative to other models have been performed on a corpus of 105 metaphors between profession domains (e.g., "*A SURGEON is a BUTCHER*"), where these domains contain an average of 120 predications each (on average, 70 of these are attributional, coding taxonomic position and descriptive properties).

Aspect	Optimal-SME	Greedy -SME	ACME	Sapper *(Vanilla)*	Sapper *(Optimal)*
Avg. Number of mid-level pmaps	269 per metaphor	269	12,657	18	18
Average Run-Time per Metaphor	N/A - worst case $O(2^{269})$ seconds	17* Seconds	N/A in time-frame	12.5• seconds	720• seconds

** Running on a 166 MHz Pentium* • Running on a SPARC 2

Table I. Comparaitive Run-Time Evaluation of SME and ACME and Sapper.

Sapper's long-term memory for these profession domains is coded via a semantic network of 300+ nodes with just over 1,600 inter-concept relations. Table I shows that Sapper performs better than other classical models in these domains (SME and ACME return no results for many examples in an extended time-frame, though Greedy-SME fares much better), three caveats should be stated to qualify these results. Firstly, although the average pmap measurement for Optimal-SME is clearly quite poor (inasmuch as it over-complicates the interpretation process immensely), it does underestimate its adequacy on some individual metaphors; as Ferguson (1997) has noted, Optimal-SME can map some metaphors with smaller pmap sets, e.g., *HACKER AS SCULPTOR* from 49 pmaps in 1,077 seconds, *ACCOUNTANT AS SCULPTOR* from 43 pmaps in 251 seconds, and *BUTCHER AS SCULPTOR* from 47 pmaps in 443 seconds. Second, other models can do better if they use tailored re-representations of Sapper's domains (in which, for example, attributions are ignored), but this raises problems as to the theoretical import of such re-representations. Third, these results establish whether the tested models can find some interpretation for a given metaphor but they say nothing about quality of the analogy returned.

For each test metaphor, there is an optimal set of cross-domain matches, so to assess the quality of a given interpretation, one needs to note how many of the produced matches actually intersect with this optimal set (as generated by the exhaustive variant of Sapper profiled in Table I), taking into account the number of "ghost mappings" (i.e., matches included in the interpretation that should not have been generated).

Table II shows some quality results for the more efficient structure mappers, Vanilla Sapper and Greedy-SME (Greedy-SIM is our simulation of Greedy-SME earlier reported in Veale & Keane, 1997, and Greedy-SME is based on an analysis of the outputs provided to us by the SME Group). Three measures of quality are used (borrowing some terms from the field of information retrieval, e.g., Van Rijsbergen, 1979). *Recall* is the total number of optimal mappings generated measured as a percentage of the total number of optimal mappings available. *Precision* is the number of optimal mappings generated measured as a percentage of the total number of optimal mappings generated by the model. Recall indicates the productivity (or under-productivity) of a model, while precision indicates over-productivity (or the propensity to generate "ghost mappings"). Finally, we measured the percentage of times a perfect, optimal interpretation was produced by the model.

The results shown in Tables I and II lead one to conclude that while Sapper and Greedy-

---

[1] Ferguson et al. (1997) have argued that Sapper cannot exploit matches based on extended chains of relational links. While many of the chains are quite short in the professions domains, recent tests have shown that Sapper has no difficulties with longer chains.

Aspekt	Vanilla Sapper	Greedy SIM	Greedy SME
Merge Complexity	$O(\Vert \log_2(\Vert)+\Vert)$	$O(\Vert \log_2(\Vert)+\Vert)$	$O(\Vert \log_2(\Vert)+\Vert)$
Precision	95%	56%	60%
Resall	95%	72%	72%
% of Times Optimal	77%	0%	0%

Table II. Quality of inetrpretation Generated by of Sapper, Greedy - SIM and Greedy - SME

SME take roughly the same time to process metaphors, the quality of the latter lags behind the former. Our analyses suggest that the specific features underlying the proposed principles contribute to Sapper's better performance, namely: its pre-preparation of potential mappings in memory, the use of a richly elaborated semantic memory and its exploitation of low-level similarity (the final issue to which we now turn).

ATTRIBUTES ARE IMPORTANT

The third main difference in principle that emerges from Sapper is its emphasis on attribute knowledge (also a cornerstone of the FARG models of Hofstadter *et al.*, 1995). For Sapper, attribute knowledge is always *necessary* to ground the mapping process, whereas in non-classical models it tends to be merely *sufficient*.

A central tenet of structure mapping theory (see Gentner, 1983) is that analogy rests on relational rather than attribute mappings, although the, sometimes misleading, influence of attribute mappings have been well-recognised (Gentner, Ratterman & Forbus, 1993; Gentner & Toupin, 1986; Keane, 1985; Markman & Gentner, 1993). Originally, in Optimal-SME, analogies were found using analogy match-rules which explicitly ignored attribute correspondences (unless they the arguments to relational matches; see Falkenhainer et al., 1989) and literally-similar comparisons were handled by literal-similarity rules that matched both relations and attributes. More recently, SME uses literal-similarity rules for both analogies and literally-similar comparisons (see e.g., Markman & Gentner, 1993; Forbus, Gentner & Law, 1995). So, if a comparison yields mainly systematic relational matches then it is an analogy, whereas if it yields more attribute than relational matches then it is literally similar. However, even though literal-similarity rules are used, attribute information is typically only sufficient in the formation of analogies, rather than necessary. If attribute matches are absent then SME will find a systematic relational interpretation for the two domains, and if they are present then it will find the same systematic relational interpretation *along* with any consistent attribute matches[2].

In contrast, Sapper proposes a strong causal role for the grounding of high-level correspondences in initial attribute correspondences. This model will simply not find any matches unless they are, in some way, grounded in attribute knowledge. The triangulation rule establishes a candidate set of mappings using category information that anchors the later construction of the analogy, so that correspondences established by the squaring rule are built on the bridges found by the triangulation rule. Thus, Sapper assumes that categories exist to enable people to infer shared causal properties among objects.

There are several psychological and computational observations that support this emphasis on the importance of attributes. First, as we already know, human memory has a tendency to retrieve analogues with have attribute overlap (see e.g., Keane, 1987; Gentner, Ratterman & Forbus, 1993; Holyoak & Koh, 1987), which must mean that many everyday analolgies rely heavily on attribute overlaps (unlike the *attribute-lite* analogies used to illustrate most analogies, like the atom/solar system and heat-flow/water-flow examples).

---

[2] This not to deny that attribute matches *can* be necessary to finding an analogy. For example, if there are two competing relational interpretations with equal systematicity, then attribute matches could tip the balance in favour of one. Similarly, in cross-mappings, attribute matches can misdirect the comparison process (cf. Markman & Gentner, 1993). However, our intuition is that these situations are the exception rather than the rule

Second, category information constrains the computational exercise of finding a structure mapping. When reasoning about two analogical situations, people will intuitively seek to map elements within categories; for instance, when mapping Irangate to Watergate, presidents will map to presidents, patsies to patsies, reporters to reporters, and so on. With these initial, tentative mappings in mind, the structure-mapping exercise that follows may be greatly curtailed in its combinatorial scope (for supporting psychological evidence see Goldstone & Medin, 1994; Ratcliff & McKoon, 1989).

Third, the triangulation of attributive information allows Sapper to model an important aspect of metaphor interpretation that has largely been ignored in most classical structure-mapping models, namely *domain incongruence* (Ortony, 1979; Tourangeau & Sternberg, 1981). The same attribute can possess different meanings in different domains and this plurality of meaning serves to ground a metaphor between these domains. For instance, when one claims that a "tie is too loud", the attribute LOUD is being used in an acoustic and a visual sense; a GARISH tie is one whose colours invoke a visual counterpart of the physical unease associated with loud, clamorous noises. But for LOUD to be seen as a metaphor for GARISH such attributes must possess an internal semantic structure to facilitate the mapping between both. That is, attributes may possess attributes on their own (e.g., both LOUD and GARISH may be associated with SENSORY, INTENSE and UNCOMFORTABLE). The division between structure and attribution is not as clean a break then as classical models predict; rather structure blends into attribution and both should be handled homogenously. This homogeniety is perhaps one of the strongest features of non-classical models.

This asserted centrality of attribute information in the mapping process may seem to be contradicted by evidence of aptness ratings on analogy, which show that apt analogies have few attribute overlaps (see Gentner and Clement, 1988; also soundness, see Gentner, Ratterman & Forbus, 1993)[3]. However, there is a possibility that these ratings may just reflect a folk theory of analogy. More plausibly, since we argue that the role of attributes is to ground high-level structure in low-level preception, the effect of this grounding may not be apparent to subjects, particularly when this grounding occurs at a significant recursive remove (e.g., H = 5). Ultimately then, these aptness ratings may tell us nothing about what actually facilitates the process of structural mapping.

CONCLUSIONS

In this paper, we have tried to show that a very different computational treatment of structure mapping in a localist semantic-memory diverges from so-called classical models of analogy in three important respects. Models like Sapper promote the idea that memory is continuously laying the groundwork for analogy formation, that analogical mapping should be memory-situated, and that attribute correspondences play a key role in the mapping process. Computationally, it is clear that at least one instantiation of these ideas does a very good job at dealing with the computational intractability of structure mapping, albeit in a sub-optimal fashion. Our experiments, both on our own profession domain metaphors (in which Sapper out-performs other models) and the benchmark analogies of other models (such as KARLA AS ZERDIA and SOCRATES AS MIDWIFE, where Sapper does at least as well as SME and ACME), suggest that of all the attempts at sub-optimal mappings it seems to offer the best all-round performance. Psychologically, much needs to be established to determine if these ideas are indeed the case. It clearly presents an interesting a fruitful direction for future research.

---

[3] As a contrasting view, note that Tourangeau & Sternberg (1981) argue that aptness is based on attribution and domain incongruence.

To conclude, should readers wish to examine the experimental data used in this research, it can be obtained (in Sapper, SME and ACME formats) from the first author's web-site: *http://www.compapp.dcu.ie/~tonyv/metaphor.html* A Prolog implementation of the Sapper model is also available from this location.

REFERENCES

T. C. Eskridge. (1994). A hybrid model of continuous analogical reasoning. In Branden (ed.), *Advances in Connectionist and Neural Computation Theory*. Norwood, NJ: Ablex.

B. Falkenhainer, K. D. Forbus, & D. Gentner. (1989). Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence*, 41, 1-63.

R. Ferguson. (1997). *Personal Communication*.

R. Ferguson, K. D. Forbus & D. Gentner. (1997). On the proper treatment of noun-noun metaphor: A critique of the Sapper model. *Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Society*. NJ: Erlbaum.

K. D. Forbus, R. Ferguson & D. Gentner. (1994). Incremental Structure-Mapping. *Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society*. NJ: Erlbaum.

K. D. Forbus, D. Gentner & K. Law. (1995). MAC/FAC:A model of similarity-based retrieval. *Cognitive Science*, 19, 141-205.

K. D. Forbus D. & D. Oblinger (1990). Making SME pragmatic and greedy. *Proceedings of the Twelfth Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: LEA.

D. Gentner. (1983). Structure-Mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.

D. Gentner & C. Clement. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. *The Psychology of Learning & Motivation*, 22. New York: Academic Press.

D. Gentner, M.J. Rattermann, & K. Forbus. (1993). The roles of similarity in transfer: Separating retrievability From inferential soundness. *Cognitive Psychology*, 25.

D. Gentner & C. Toupin. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, 10, 277-300.

R.L. Goldstone & D.L. Medin. (1994). Time course of comparison. *Journal of Experimental Psychology: Language, Memory & Cognition*, 20, 29-50.

D. R. Hofstadter & the Fluid Analogy Research Group (1995). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. Basic Books, NY.

D. R. Hofstadter & R. French (1995). The Table-Top system: Perception as Low-Level Analogy, in *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought (ed. D. Hofstadter), chapter* 9. Basic Books, NY.

K.J. Holyoak & K. Koh (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, 15, 332-340.

K. J. Holyoak & P. Thagard. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295-355.

J. E. Hummel & K. J. Holyoak. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*.

M. Keane (1985). On drawing analogies when solving problems: A theory and test of solution generation in an analogical problem solving task. *British Journal of Psychology*, *76*, 449-458.

M.T. Keane (1987). On retrieving analogues when solving problems.*Quarterly Journal of Experimental Psychology*, *39A* , 29-41.

M. T. Keane & M. Brayshaw. (1988). The Incremental Analogical Machine: A computational model of analogy. *In D. Sleeman (Ed.), European Working Session on Learning*. Pitman, 1988.

M. T. Keane, T. Ledgeway & S. Duff. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science*, 18, 387-438.

B. N. Kokinov (1994). A hybid model of reasoning by analogy. In K.J. Holyoak & J.A. Barnden (Eds.) *Advances in Connectionist and Neural Computation.* Norwood, NJ: Ablex.

A. Markman & D. Gentner. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25, 431-467.

A. Ortony. (1979). The role of similarity in similes and metaphors. In A. Ortony (Ed.) *Metaphor and Thought.* Cambridge, MA: Cambridge University Press.

R. Ratcliffe & G. McKoon (1989). Similarity information versus relationla information. *Cognitive Psychology*, 21, 139-155.

R. Tourangeau & R.J. Sternberg (1981). Aptness in metaphor. *Cognitive Psychology*, 13, 27-55.

P. Thagard. (1997). *Personal Communication.*

C. J. van Rijsbergen. (1979). *Information Retrieval.* Butterworths.

T. Veale & M. T. Keane. (1993). A connectionist model of semantic memory for metaphor interpretation. *Workshop on Neural Architectures and Distributed AI*, 19-20, the Center for Neural Engineering, U.S.C. California.

T. Veale & M. T. Keane. (1994). Belief modelling, intentionality and perlocution in metaphor comprehension. *Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

T. Veale & M. T. Keane. (1997). The competence of sub-optimal structure mapping on 'hard' analogies. *IJCAI'97: The 15th International Joint Conference on A.I. Morgan Kaufmann.*

T. Veale, D. O'Donoghue & M. T. Keane. (1996). Computability as a limiting cognitive constraint: Complexity concerns in metaphor comprehension, *Cognitive Linguistics: Cultural, Psychological and Typological Issues (forthcoming).*

Appendix A: Pseudocode of the Sapper Algorithm

```
Function Sapper::Stage-I (T:S, H)
    Let
    Spread Activation from roots T and S in long-term memory to a horizon H
When a wave of activation from T meets a wave from S at a bridge T':S' linking a target
    domain concept T' to a source concept S' then:
        Determine a chain of relations R that links T' to T and S' to S
        If R is found, then the bridge T':S' is balanced relative to T:S, so do:
            Generate a partial interpretation p of the metaphor T:S as follows:
        For every tenor concept t between T' and T as linked by R do
    Align t with the equivalent concept s between S' and S
    Let {t:s}
Let {p}
Return P, a set of intermediate-level pmaps for the metaphor T:S

Function Sapper::Stage-II (T:S, P)
Once all partial interpretations $P = \{p_i\}$ have been gathered, do:
    Evaluate the quality (e.g., mapping richness) of each interpretation $p_i$
        Sort all partial interpretations $\{p_i\}$ in descending order of quality.
        Choose the first interpretation G as a seed for overall interpretation.
            Work through every other pmap $p_i$ in descending order of quality:
            If it is coherent to merge $p_i$ with G (i.e., respecting 1-to-1ness) then:
            Let $_i$
        Otherwise discard $p_i$
    When {p} is exhausted, Return G, the Sapper interpretation of T:S
```

AN ARGUMENT FOR DERIVATIONAL ANALOGY

Erica Melis
Universitaet des Saarlandes, Fachbereich Informatik
D-66041 Saarbruecken, Germany, melis@ags.uni-sb.de

Jaime G. Carbonell
Carnegie Mellon University, School of Computer Science
Pittsburgh PA 15213, U. S. A., jgc@cs.cmu.edu

1. INTRODUCTION

A common reason for the use of analogy in (computational) problem solving is the lack of appropriate object-level knowledge, e.g. rules, necessary to solve the problem from first principles. Hence, the absence of sufficient (object-level) domain knowledge is assumed in most case-based reasoning (CBR) systems. Even those CBR systems that combine rule-based and case-based reasoning rely on a similar assumption: if rules exist, then reason from first principles, otherwise use case-based reasoning [17,18]. That is, the use of analogy as a search control strategy by transferring control knowledge, is hardly an issue in CBR research, except in case-based planning (CBP).

As far as we know, the situation is similar in cognitive research on analogy. Why this? One reason might be that more often than not the problems chosen for cognitive experiments have single-step solutions rather than solutions with many steps as in planning and hence, search control does not matter much. For instance, the much investigated/standard problems "atom/solar system", "water flow/ heat flow", and Duncker's radiation problem do not require a search-intensive multi-step solution process.

As opposed to solutions of these problems, Newtonian physics problem solving [20] and especially mathematical theorem proving need a complex multi-step problem solving process, where search control is a central issue. The same is true for many computational planning problems. The problems to be solved by CBP may have complex and multi-step solutions, e.g., in mathematical theorem proving [12]. Therefore, CBP aims at reducing the search effort for finding a solution [5, 22, 1, 11].

This paper is centered around our experiences with problem solving for complex solutions that have multiple steps, where decisions as to which sequences of steps to explore are crucial. Here, problem solving by analogy can have the following purposes:

Computational analogy tries to improve the exploitation of limited resources, in particular of the number of user interactions, run time, and of knowledge. Hence, the purpose of analogy can be, cf. [10] to save user interaction (which is a replacement for control knowledge in interactive systems); to use analogy to replace search-intensive subroutines at low cost.

Similarly, for *human problem solving by analogy* Van Lehn and Jones [20] suggest that at least good human problem solvers use analogical problem solving:

- when no general (object-level) knowledge physics principles such as the force law, Newton's law, and mathematical transformations for solving a current problem is available, e.g., if a knowledge gap has to be detected and filled; For instance, subjects detected a force that was missing in a diagram by checking a previous solution. Detecting a gap means to discover that some principle is missing for a problem to be solved.
- when specific information from an example can be used in order to work more efficiently D in other words to save

search., e.g., for the explication of physics quantities.

Put differently, the computational experience and the described cognitive results suggest that the task of *analogical transfer of multi-step problem solving*, requires to (1) transfer object-level knowledge *and/or* (2) control knowledge, that is, decisions on the choice of steps, instantiations, etc. As for the second, the decisions may well depend on the problem solving *context*. Therefore, the transfer of control knowledge requires:

- to *check whether the target context justifies a decision* as the source context did. This check has to be performed immediately before each step transfer because each step in a solution process builds on results of earlier steps and hence a whole transferred solution may be invalidated by the failure of an intermediate step and a simple modification of this failed step alone cannot guarantee to yield a valid solution;
- to actually *replay source decisions* in the target. These decisions may differ considerably from the actual solution steps, e.g., the decisions may concern abstract steps that can yield different results when executed in different situations.

1.1 Contribution of the Paper

As explained in §2, *derivational analogy* is a computational answer to the described needs of transferring control knowledge in analogical problem solving. In that section we discuss our experiences with derivational analogy in a transportation planning domain and in mathematical proof planning. Furthermore, section 2.2 explains the transfer of object-level knowledge by reformulation that can be combined with derivational analogy. Section 2.2.1 discusses some advantages of derivational analogy compared to the pure transformational approach assumed in most cognitive models.

Then we address the question whether computational derivational analogy can model human analogical transfer of multi-step solutions under certain conditions. We suggest some questions to be addressed empirically, e.g., 'Do characteristics from computational derivational analogy transfer to the spontaneous or guided use of analogical problem solving?' In particular, we suggest questions whose empirical answer can contribute to a well-founded *support* of analogical problem solving, say in teaching and assistant systems.

Our expertise is in computational analogy. Therefore our questions and suggestions should be considered a mere proposal for further cognitive and multidisciplinary research.

2. DERIVATIONAL ANALOGY

Derivational analogy introduced in [6] denotes a process that draws analogies from the experiences of the past reasoning *process*. The underlying key insight is that useful experience is encoded in the reasoning process used to derive solutions to similar problems, rather than just in the final solution. Therefore, derivational analogy is a *reconstructive* method by which lines of reasoning, i.e., of search control, are transferred and adapted to a new problem as opposed to transformational analogy that adapts the final solutions.

The derivational analogy framework has been instantiated by several computational systems, including BOGART [14], REMAID [3], PRIAR [7], APU [2], Prodigy/Analogy [23], and ABALONE [13]. These systems apply to a variety of multi-step problem solving activities, including software reuse in a UNIX programming domain [2], the design of human computer interfaces [3], and several planning applications [7, 22, 13].

The case-based planning is built on top of a generative planner that generates the source plans consisting of operators reducing a goal to subgoals. Typically, this generative planning involves a lot of search because several operators are applicable to a goal. Case-based planning *suggests* the choice of operators rather than *searching* for them.

If possible, the derivational analogy replays the choice of operators in a source plan step by step. If the justification of a particular choice

```
input: source plan, source and target problem
output: (partial) target plan Map source and target.
Map source and target.
while source plan not exhausted do
    Get next operator M from source plan.
    Check M's justifications.
    if justifications hold, then transfer M to target
      and advance source,
    else choose suitable action.
Base-level plan for remaining open goals.
```

Table 1. Top-Level Algorithm of Derivational CBP.

does not hold in the target, then it may be possible to carry out some reaction. As the outline in Table 1 shows, the implementations of derivational case-based planning have three main components, the retrieval including the mapping from source to target, the check of justifications for a source decision in the target and the actual replay in case the justification holds or can be established. This analogy permits a partial transfer of solutions when a total transfer cannot be justified.

In order to check of justifications during the analogical replay, these justifications have to be stored and indexed. Automatic generation of the derivational planning episodes occurs by extending the base-level generative planner with the ability to examine its internal decision cycle, recording the justifications, i.e., reasons why an operator was chosen, for each decision during its search process. Veloso [22] discusses the importance of choosing relevant justifications and of providing a language for justifications: The stored information should be directly available during the generative planning and relevant information should be stored only.

2.1 Analogy in Complex Planning Domains

Planning systems in Artificial Intelligence fall into two general categories:

1. Hierarchical "top-down" planners, such as SIPE [25], which can solve relatively complex problems but require significant knowledge engineering of each new domain, and also exhibit somewhat rigid planning behavior.
2. Operator-based "bottom-up" planners, such as PRODIGY [24], which often require massive search to solve complex problems, but make do with simpler knowledge engineering and exhibit more robust behavior, including the production of different contingency plans.

In order to combine the best features of both paradigms, the Prodigy project has integrated non-linear operator-based planning with multiple types of learning, including control-rule learning, representation-change learning, abstraction-hierarchy learning, and derivational analogy. Learning provides search guidance and makes more complex problems tractable, while retaining the underlying flexibility of the operator-based planner if necessary Ð i.e. when previously acquird knowledge proves insufficient in solving a novel problem. Derivational analogy has proven particularly useful in this regard [23].

Among several application domains, Prodigy was used to produce plans that solve transportation/logistics problems whose solution may require several hundreds individual steps. The transportation domain involves moving multiple sets of objects through an inter-city transportation network relying on different vehicles (trucks, airplanes), with preference for lower-cost solutions [26]. Prior to attempting complex problems, Prodigy was trained with simple problems, then increasingly more complex ones, which led to the creation of a 1000-case library [22]. Rather than delving into the details previously reported in the literature, let us focus on the lessons learned:

- *Control Knowledge is Crucial Ð* In theory, all well-defined transportation problems can be solved or proven unsolvable by the first-principles planner. But, in practice, base-line Prodigy would require search spaces several orders of magnitude larger than its maximum capabilities to solve 200-step non-linearly-decomposable transportation problems. Hence, Analogy expands the solvability horizon of a planner just by supplying much-needed control knowledge.

- *Reasoning with Justifications is Crucial* - Pure transformational CBR does not check justifications. These are crucial, however, to guarantee the soundness of the retrieved analog plans for the current problem - or to repair the plan if the justifications fail. Derivational analogy works because all plans are equally reliable Ð there is no tradeoff between careful reasoning and risky memory lookup, as justification checking eliminates the risk of inapplicability.

- *Interleaving Analogical Rederivation with First-Principles Planning is Crucial* - Many complex problems can be partially but not fully solved by rederivation of past cases. For instance, a particular road used before may be closed, or all airplanes in a particular city may be grounded by fog. Or, simply, the problem places some new demands not previously encountered. Justification failure is an invitation to reason from first principles either to re-establish the failed justification (e.g. wait for the fog to clear), or to keep the bulk of the plan and modify the failed part (e.g. keep the same route, but detour around the closed segment).

- *Interleaving Multiple Cases is Very Useful* - Most often, past cases solve parts of the new problem, and several must be composed, with occasional gaps filled in by first-principles planning, in order to solve increasingly complex problems.

- *Derivational Analogy Does Not Sacrifice Plan Efficiency* - Derivational analogy plans efficiently, but does it produce efficient plans? This is a legitimate question best answered empirically, since neither first-principles nor analogical planning guarantees optimality. Test showed equivalent plan execution cost on average for the transportation domain. Explicit learning of plan-efficiency control rules, however can help both base-level and derivational analogy planning produce plans that minimize execution cost [26].

- *Knowledge Revision is an Unresolved Issue* - An unresolved issue is how to modify a large analogical case library if the domain knowledge changes significantly. For instance, if a new mode of transportation is invented replacing trucks (as the latter replaced horse-drawn carts), past plans become obsolete. However, if smaller, more subtle changes occur (e.g. a new speed limit is enacted), it should prove feasible to salvage the plan library. Whereas this issue remains unresolved, some domains such as theorem proving (discussed below) need not worry about the underlying mathematical knowledge ever reaching obsolescence.

2.2 Analogy in Planning Proofs of Mathematical Theorems

Proof planning is a methodology for automated theorem proving that constructs a proof by search at the abstract level of proof plans [4]. On top of a proof planner, analogy-driven proof plan construction [9] yields a (partial) proof plan that may be expanded to a proof. Analogy-driven proof plan construction is an extension of the general derivational CBP because it extends the mapping to a second-order mapping, new kinds of justifications, described below, extend those in simpler planning domains, and because it includes *reformulations* of the source plan as shown in Table 2.

Sometimes a step by step replay will not be enough. In this case, the source plan may be reformulated before the replay. Reformulations can insert, change, or delete source operators. They map proof plans in a way based on differences between the source and target problems, i.e., they are triggered by peculiarities in the second-order mapping. For instance, the reformulation 1to2 is triggered when there is a C-equation $f_i=f_j$ (see below) and the mapping m_e from source to target violates the equation as follows me(fi)(me(fi))=me(fj) . 1to2 changes a one-step induction in the source to a two-step in the target. In addition, it doubles certain operators in the target plan.

```
input: source plan, source theorem and assumptions, target theorem and assumptions
output: (partial) target plan
Second-order map source and target triggers reformulations of the source plan.
source plan ← reformulated source plan.
    while source plan not exhausted do
    Get next operator M from source plan.
    Check M's justifications.
    if justifications hold, then transfer M to target
      and advance source,
else choose suitable action.
Plan from first principles for remaining open goals.
```

Table 2. Outline of Analogy-Driven Proof Plan Construction.

Since an operator such as induction computes its outputs, the actual subproof that is represented by the operator may vary between solution. Hence operators are abstract entities in the solution and an analogical replay requires to actually apply a chosen operator in the target in order to get the correct output.

Now we present some justifications and explain the reaction to failed justifications with the following example where the source problem is a theorem and lemmata about lists.

The source proof plan and has operators such as induction, elementary for trivial subproofs, and wave which we won't explain here. (Note, however, that operators such as induction or elementary may produce different subproofs in different situations.) The target problem is one about natural numbers:

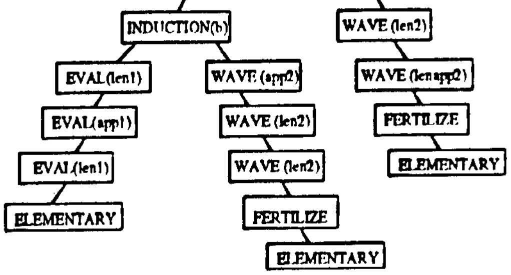


Figure 1. Proof Plan of the Theorem lennapp.

Source Theorem (lennapp):
$length(app(a,b) = length(app(b,a))$
Lemmata:
app2: $app(cons(X,Y),Z) \Rightarrow cons(X, app(Y,Z))$
len2: $length(cons(X,Y)) \Rightarrow s(length(Y))$
lenapp2: $length(app(X, cons(Y,Z))) \Rightarrow s(length(app(X,Z)))$

Target Theorem (halfplus)
$half(+(a,b)) = half(+(b,a))$
Lemmata:
(plus2): $+(s(Y),Z) \Rightarrow s(+(Y,Z))$
(half3): $half(s(s(Y))) \Rightarrow s(half(Y))$.

Justifications Stored in the Source Plan

The analogy system ABALONE is implemented on top of the proof planner CL^M that stores two new kinds of justifications:

- Legal conditions on the context for the application of operators, such as the existence of a lemma that is necessary to apply the wave operator.
- Constraints on the objects (e.g., the function symbols) that are required for the source solution, in particular the identity of different occurrences of a function symbol.

Since ABALONE is able to send a function symbol at different positions in the source to different target images, source function symbols at different positions are differentiated by indices. Then the source problem becomes (only some indices are shown for simplicity).

Source Theorem (`lennapp`):
$length_1(app_1(a,b) = length_2(app_2(b,a))$
`Lemmata`:
`app2`: $app(cons_1(X,Y),Z) => cons_2(X, app(Y,Z))$
`len2`: $length_3(cons_3(X,Y)) => s_1(length_4(Y))$
`lenapp2`: $length_5(app(X, cons_4(Y,Z))) => s_2(length_6(app(X,Z)))$

During the source planning, constraints may be placed on these indices, yielding *C(onstraint)-equations*, of the form fi = fj in the source plan. These C-equations form an additional justification that must be satisfied in the target for a successful replay. The following C-equations emerge from the source planning process for the source problem lenapp: cons5=cons1 , cons2=cons3 , cons 5 = cons 4, s1=s2 , where cons5 is introduced by the induction operator.

We have to consider the mappings found for the example in order to understand how they violate justifications. The second-order basic mapping mb for the theorems is: lengthi half , and appi + (for all i). mb is extended to a mapping me that maps the source and target lemmata. For instance, lemma app2 is mapped in the following way:

app(cons1(X,Y), Z) ⇒app(cons2(X,Y),Z) (app2)¯mb: <>+(cons1(X,Y),Z) ⇒ cons2(X, +(Y,Z)) +(s(Y),Z) ⇒s(+(Y,Z)) (plus2) me (cons1,2) = lw1.lw2. s(w2)

Since mb(app)=+ , lemma app2 can be partially mapped to +(cons1(X,Y),Z)⇒cons2(X, +(Y,Z)) . The mapping is completed by mapping the source lemma to an available target lemma. In this way, app2 maps to plus2 with cons1 lw1lw2 . s(w2) and cons2 lw1lw2 . s(w2) . Similarly, len2 maps to half3 because of mb(length) = half , giving me(s1)=lw1. s(w1) and me(cons3)= lw1.w2. s(s(w2)) . Note that the latter violates the C-equation cons3=cons2 , because cons2, cons3 have different target images.

2.2.1 Reaction to Failed Justifications

If the check of a justification fails during ABALONE's analogical replay, certain reactions to failed justifications try to make an operator applicable anyway, for instance:

1. If a justification that requires the existence of a certain lemma does not hold in the target, i.e., if a target lemma corresponding to a certain source lemma cannot be found, then ABALONE speculates a target lemma.
2. If a C-equation is violated, then a reformulation is applied under certain conditions.

In the example a violation of the C-equation cons3=cons2 occurs because me(cons3)= me(cons2) and this triggers a 1to2 reformulation which duplicates the operator wave(app2) such that the resulting target plan contains two operators wave(plus2).

The first kind of failure occurs in the example since the source lemma lenapp2 does not have an image in the target because it cannot be mapped to plus2 or half3 by extending mb . The appropriate reaction is to speculate a target lemma. ABALONE uses the mappings and C-equations s2=s1 with the mapping me(s1)=lw1 .s(w1) s(w1) , and cons4=cons5 with the mapping me(cons5)=lw1lw2 . s(s(w2)) to come up with the target lemma:

half(+(X, s(s(Z)))) ⇒s(half(+(X,Z))) as an image oflength(app(x, cons4 (y,z))) ⇒s2 (length(app(x,z))) .

2.2.2 Summary

Derivational analogy is needed because the replaying an (abstract) decision in a certain situation may result in a concrete solution that cannot be obtained by simply transferring steps (e.g., different logical proofs produced by running the elementary operator in different situations).

Justifications are crucial since they can they guarantee the soundness of steps chosen by analogy for a target problem.

Reasoning about justifications is crucial because this allows to derive reactions to failing justifications in the target, even depending on the available resources.

Justifications may serve as explanations in proofs presented to a user.

2.2.3 Advantages of Derivational Analogy

Carbonell [6] discusses an example illustrates an advantage of derivational analogy: Suppose you have coded a quicksort routine in Pascal, and then you are asked to recode the routine in LISP. Although the problem-solving process may preserve much of the inherent similarity, the resultant solutions may be hardly similar. A line-by-line program transfer is clearly not appropriate, but a reuse of major structural and control decisions required to construct the Pascal program is possible. Therefore, the analogy must be guided by a reconsideration of the key decisions in light of the new situation. In particular, the derivation of the LISP quicksort program starts from the same specification, retaining the same divide-and-conquer strategy, but it may diverge in the selection of data structures (list vs.arrays) or in the method of choosing the comparison element. However, future decisions that do not depend on earlier divergent decisions can still be tran!!sferred to the new domain rather than recomputed.

Similarly, in proof planning several operators represent an *abstraction* of the actual subproof they produce. For instance, an application of the operator induction involves computing an induction schema, the induction variables, the base case and the step case subgoal of the proof. For instance, elementary can produce different proofs when executed. Thus the replay of a proof plan in different situations can result in different proofs and different subgoals although abstractly the source proof equals the target proofs.

From the above examples other advantages of derivational analogy can be summarized.

3. INTERESTING QUESTIONS

The above description of analogical search control suggests the question 'How does all this apply to human analogical problem solving?' which implies many more specific questions to cognitive research:

1. Can justifications/derivational information be found in spontaneous human analogy?
2. Is storing derivational information psychologically implausible because of the limited working memory as proposed in [16] ? Is it necessary, as suggested by Reimann, to store as much as possible from a problem solving episode?
3. What are relevant justifications in human problem solving? Are they domain-dependent?
4. Does memorizing *relevant* justifications depend on expertise and on the ability of self-explanation.
5. How do expert self-explanation and extracting justifications from a problem solving process relate?
6. What is the impact of carefully chosen derivational information on analogical transfer and adaptation performance?
7. Can adaptation schemas be found in human analogical problem solving? How do they compare with reformulations triggered by failed justifications?
8. Can context, as addressed in [8], be modelled by derivational information?
9. How do explanations as addressed in [15] compare with justifications?
10. Which experimental techniques can (nearly) exclude mental reference to derivational information that cannot be observed as opposed to explicit reference?
11. For research that cares about supporting analogical reasoning, for instance for tutor systems, the following questions may be particularly interesting.
12. What is the influence of externally provided derivational information on performance and correctness of human analogical problem solving? Hence, which information should be provided in teaching and tutor systems to support the analogical problem solving?
13. Does derivational information support people in noticing analogies?
14. Does derivational information create self-explanations?
15. Does derivational information support learning from analogies?

3.1 Related Work

Van Lehn [19] suggests that a solver who 'understands how an example's result is derived can adapt it more intelligently to the target problem. Thus, one would expect the Good solvers to use derivational analogy more frequently than non-derivational analogy and Poor solvers should use non-derivational analogy more than derivational analogy.' To check this prediction, Van Lehn analyzed transfer events in Newtonian physics learning to see if the student explained the example before transferring it. VanLehn concludes that self-explaining the example during analogical problem solving is not particularly common. We think that **the experiments could be varied, however, by providing (written) explanations in the source problem solving and by experimenting with more difficult multi-step solutions where derivational analogy might be necessary.**

Van Lehn models some analogical search control in Cascade [21]. It stores triples consisting of the problem, the goal, and the rule to achieve the goal. Whenever faced with a search control decision, Cascade decides by analogy. Thereby Cascade could learn rules that it could not otherwise learn. Analogical search control modeled the intuition that students learn more than just physics rules from studying the examples because they also learn "how to" knowledge. In experiments Cascade's analogical search control did not match well with the protocols. In the opposite, a default ordering of rules plus few general search heuristics did sufficiently explain the subject's behavior. We think that **(1) the latter explation should be checked with more complicated solutions for which rating the steps is far from sufficient, e.g., in proof planning. (2) Instead of always deciding by analogy, we would expect analogical search control only in case the search space is la!!rge, i.e., many alternative decisions are possible.**

Reimann [16] discusses that derivational analogy is a normative model for high-quality analogical problem solving. He thinks of it as implausible though.

4. CONCLUSIONS

Based on our experience in computational analogy, we pointed to characteristics and advantages of derivational analogy in problem solving. We discuss case-based planning for problems of the transportation domain and of mathematical proof planning. As opposed to transformational analogy, derivational analogy provides analogical search control based on justifications for decisions. The choice and design of the justifications is of great importance to the computational analogy systems. Does this hold for human solvers too?

The derived questions and suggestions propose further cognitive and multidisciplinay research, in particular, for supporting analogical reasoning on complex problems. Vice versa, cognitive empirical results are essential in order to acquire and represent the right knowledge in computational systems that are supposed to model or to support human analogical problem solving, e.g. in a proof planner.

REFERENCES

[1] R. Bergmann and W. Wilke. On the role of abstraction in case-based reasoning. In B. Faltings and I. Smith, editors, *Fourth European Workshop on Case-Based Reasoning (EWCBR-96),* Lausanne, 1996.

[2] S. Bhansali and M.T. Harandi. Synthesis of UNIX programs using derivational analogy. *Machine Learning,* 10, 1993.

[3] B. Blumenthal. *Replaying Episodes of a Metaphor Application Interface Designer.* PhD thesis, University of Texas, Artificial Intelligence Lab, Austin, 1990.

[4] A. Bundy. The use of explicit plans to guide inductive proofs. In E. Lusk and R. Overbeek, editors, *Proc. 9th International Conference on Automated Deduction (CADE-9),* volume 310 of *Lecture Notes in Computer Science,* pages 111-120, Argonne, 1988. Springer.

[5] J.G. Carbonell. Learning by Analogy: Formulating and Generalizing Plans from Past Experience. In R.S. Michalsky, J.G. Carbonell, and T.M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach I.* Morgan Kaufmann Publ. Los Altos, 1986.

[6] J.G. Carbonell. Derivational analogy: A theory of reconstructive problem solving and expertise aquisition. In R.S. Michalsky, J.G. Carbonell, and T.M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach II,* pages 371-392. Morgan Kaufmann Publ. Los Altos, 1986.

[7] S. Kambhampati and J.A. Hendler. A validation based theory of plan modification and reuse. *Artificial Intelligence,* 55:193-253, 1992.

[8] B.N. Kokinov. A hybrid model of reasoning by analogy. In K.J. Holyoak and J.A. Barnden, editors, *Analogical Connections and Neural Computation Theory,* pages 247-320. Ablex, 1994.

[9] E. Melis. A model of analogy-driven proof-plan construction. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence,* pages 182-189, Montreal, 1995.

[10] E. Melis When to prove theorems by analogy? In *KI-96: Advances in Atrificial Intelligence. 20th Annual German Conference on Artificial Intelligence,* volume 1137 of *LNAI,* pages 259-271. Springer, 1996.

[11] E. Melis and J. Whittle. Internal analogy in inductive theorem proving. In M.A. McRobbie and J.K. Slaney, editors, *Proceedings of the 13th Conference on Automated Deduction (CADE-96),* Lecture Notes in Artificial Intelligence, 1104, pages 92-105, Berlin, New York 1996. Springer. also published as DAI Research Paper 803.

[12] E. Melis and J. Whittle. Analogy as a control strategy in theorem proving. In *Proceedings of the 10th Florida International AI Conference (FLAIRS-97),* pages 367-371, 1997. also published as DAI Research Paper 840, University of Edinburgh, Dept. of AI.

[13] E. Melis and J. Whittle. Analogy in inductive theorem proving. *Journal of Automated Reasoning,* 20(3): - 1998. to appear.

[14] J. Mostow. Design by derivational analogy: Issues in the automated replay of design plans. *Artificial Intelligence,* 40(1-3):119-184, 1989.

[15] G. Nelson, P. Thagard and S. Hardy. Integrating analogy with rules and explanations. In K.J. Holyoak, editor, *Analogical Connections,* pages 181-206, Ablex, NJ, 1994.

[16] P. Reimann. *Lernprozesse beim Wissenserwerb aus Beispielen.* Verlag Hans Huber, 1997.

[17] E.L. Rissland and D.B. Skalak. CABARET: Rule integration in a hybrid architecture. *International Journal of Man-Machine Studies,* 34:839-887, 1991.

[18] J. Surma and K. Vanhoof. Integrating rules and cases for the classification task. In *Proceedings of the First International Conference ICCBR-95,* pages 325-334, 1995.

[19] K. van Lehn. Analogy events: How examples are used during problem-solving. *Cognitive Science,* -(-):-, 1998. (in press)

[20] K. van Lehn and R.M. Jones. Better learners use analogical problem solving sparingly. In *Proceedings of the 10th International Conference on Machine Learning,* pages 338-345, Amherst, MA, 1993. Morgan Kaufmann.

[21] K. van Lehn and R.M. Jones. Learning by explaining examples to oneself: A computational model. In S. Chipman and A. Meyrowitz, editors, *Cognitive Models of Complex Learning,* pages 25-82. Kluwer Academic Publishers.

[22] M.M. Veloso. *Planning and Learning by Analogical Reasoning.* Springer, Berlin, New York, 1994.

[23] M.M. Veloso and J.G. Carbonell. Derivational analogy in PRODIGY: Automating case acquisition, storage and utilisation. *Machine Learning,* 10:249-278, 1993.

[24] Carbonell, J.G., Knoblock, C.A. and Minton, S.N. Prodigy: An Integrated Architecture for Planning and Learning. In *Architectures for Intelligence,* K. vanLehn (ed.) Lawrence Erlbaum, Hillsdale, NJ, 1990.

[25] Wilkin, D.E. *Practical Planning.* Morgan Kaufman, Los Altos, CA, 1988.

[26] Perez, A.M. and Carbonell, J.G. Control Knowledge to Improve Plan Quality. *Proceedings of the Second International Conference on AI Planning Systems* Chicago, IL, AAAI Press, pp.323-328, 1994.

STRUCTURED OPERATIONS WITH DISTRIBUTED VECTOR REPRESENTATIONS

Tony A. Plate
School of Mathematical and Computing Sciences, Victoria University of Wellington
Wellington, New Zealand
Email: tap@mcs.vuw.ac.nz

ABSTRACT

Holographic Reduced Representations (HRRs) are a method for encoding nested relational structures in fixed width vector representations. HRRs encode relational structures as vector representations in such a way that the superficial similarity of the vectors reflects both superficial and structural similarity of the relational structures. HRRs support a number of operations that could be very useful in models of analogy processing: fast estimation of superficial and structural similarity via a vector dot-product; chunking of vector representations; and finding corresponding objects in two structures.

1. INTRODUCTION

Vector representations are popular for memory models for a variety of theoretical and practical reasons. They are simple and support fast parallel processing such comparison via dot-products. They are also neurologically plausible, in that they can be stored and processed in networks of simple neuron-like processing elements, such as associative vector memories. However, their use in models of analogy processing has been limited by the widespread supposition that it is difficult or impossible to encode compositional structure in vector representations (Fodor and Pylyshyn ,1988, Ratcliff and McKoon, 1989, Thagard, Holyoak, Nelson and Gochfeld, 1990, Gentner and Markman, 1993, Forbus, Gentner and Law, 1994, Wharton *et al* 1994).

This supposition is false. Structure can be represented in vectors in a number of ways, e.g., Smolensky's (1990) tensor products, Pollack's (1990) RAAMs, Kanerva's (1996) binary spattercodes, and Plate's (1995) HRRs. This paper describes HRRs and makes a number of claims for their usefulness in models of analogy retrieval and processing:

- HRRs provide an adequate vector-based representation of structure (in contrast to feature-vector approaches, which must be complemented with a conventional symbolic representation for structure).
- Estimates of similarity that reflect both superficial and structural similarity can be computed quickly via vector dot-products. This technique shows similar abilities and limitations with respect to detecting similarities as are observed in people's ability to retrieve items from long term memory.
- Corresponding objects in two analogical structures can be found via fast but approximate vector-based techniques.
- HRRs provide an elegant implementation of chunking and "pointers" for complex, structured items stored as vectors in a content addressable memory.

2. ANALOGY PROCESSING IN PEOPLE

Analog retrieval and mapping have received a significant amount of attention in the psychological literature. Much attention has been devoted to teasing apart the differing effects of superficial and structural similarity in retrieval and mapping.

For illustrations, the following series of episodes are used in this paper. Together the episodes involve dogs (Fido, Spot and Rover), people (Jane, John and Fred), a cat (Felix) and a mouse (Mort). Members of one species are assumed to be similar to each other but not to members of other species. The "probe" episode, to which the others are compared, is "*Spot bit Jane, causing Jane to flee from Spot*". There are five other episodes, which have different combinations of types of similarity to the probe (all share predicates with the probe):

LS (Literal Similarity) "*Fido bit John, causing John to flee from Fido.*" (Has both structural and superficial similarity.)

SF (Surface features) "*John fled from Fido, causing Fido to bite John.*" (Has superficial but not structural similarity.)

CM (Cross-mapped analogy) "*Fred bit Rover, causing Rover to flee from Fred.*" (Has both structural and superficial similarity, but types of corresponding objects are switched.)

AN (Analogy) "*Mort bit Felix, causing Felix to flee from Mort.*" (Has structural but not superficial similarity).

FOR (First-order-relations only) "*Mort fled from Felix, causing Felix to bite Mort.*" (Has neither structural nor superficial similarity, other than shared predicates.)

It is generally accepted that in adults, structural similarity plays a large role in analogical mapping and conscious similarity judgements. The role of structural similarity in retrieval is less clear: some researchers argue that structural similarity usually has little effect on retrieval (Gentner, Rattermann, and Forbus, 1993) while others argue that under some circumstances, structural similarity can influence retrieval (Wharton *et al*, 1994). Others suggest that structural similarity matters only when the entities involved in the situations share superficial features (Ross, 1989). Overall, the general pattern for retrievability of items from long term memory seems to be LS > CM ≥ SF > AN ≥ FOR.

Existing computational models of human performance on analog retrieval tasks such as ARCS (Thagard *et al*, 1990), and MAC/FAC (Forbus, Gentner and Law, 1994) have explained the human retrieval data by invoking two processes. The first is a simple one based on superficial similarities. This explains much of the human performance, but cannot account for effects of structural similarity. Thus, these models require a second process that takes structural similarity into account, which involves additional complex computation. In this paper I will argue that HRRs can provide a single-stage model based on vector-matching that explains the pattern of retrieval ability observed in people.

3. VECTOR REPRESENTATIONS AND OPERATIONS

The two vector operations commonly used with vector representations are superposition (i.e., addition) and similarity (i.e., dot-product or cosine). These two vector operations, and other scalar-vector operations such as scaling and normalization, are sufficient for interesting and useful memory models. With the addition of the circular convolution operation for binding, one can encode associations in vector patterns which and thus encode structure.

3.1 Local & distributed representations

Vector representations come in two flavors: local and distributed. In some respects, localist and distributed representations are equivalent. They can be indistinguishable when features are numerous and fine-grained. Also, localist representations can be mapped to distributed ones by a simple linear map, and back by a thresholded linear map. However, a crucial difference is that the total number of possible features is limited to the vector dimensionality in localist representations, but is exponential in vector dimensionality in distributed representations. This gives distributed representations the capacity to represent combinatorial features (such as Wharton's *et al* (1994) sour-grapes feature "thing that is desired but can't be obtained and hence is denigrated") in a moderate sized vector.

What is needed is a systematic way of generating and decoding the patterns which represent combinatorial features. This is the role of the binding operation. As a binding operation, circular convolution provides a fast, systematic, and reversible way of constructing new patterns to represent combinatorial features.

3.2 Circular convolution

Circular convolution maps two real-valued *n*-dimensional vectors onto one. If **x** and **y** are *n*-dimensional vectors (subscripted 0 to *n*-1), then the elements of **z** = **x**⊗**y** are

$$z_i = \sum_{k=0}^{n-1} x_k y_{i-k}$$

where subscripts are taken modulo-*n*. and ⊗ denotes circular convolution. Circular convolution can be viewed as a compression of the outer (or tensor) product of the two vectors, as shown in Figure 1. Each of the small circles represents an element of the outer product of **x** and **y**, e.g., the middle bottom one is x_2y_1. The elements of the circular convolution of **x** and **x** are the sums of the outer product elements along the wrapped diagonal lines.

Circular convolution can be regarded as a multiplication operator for vectors and has many algebraic properties in common with scalar and matrix multiplication. It is commutative (**x**⊗**y**=**y**⊗**x**), associative (**x**⊗(**y**⊗**z**)=(**x**⊗**y**)⊗**z**), and bilinear (**x**⊗(α**y**+β**z**)=α**x**⊗**z**+β**x**⊗**z**). There is an identity vector **I** (**I**⊗**x**=**x**) and a zero vector $\bar{0}$ ($\bar{0}$⊗**x**=$\bar{0}$). Inverses $\mathbf{x}^{-1}$ exist for most vectors ($\mathbf{x}^{-1}$⊗**x**=**I**).

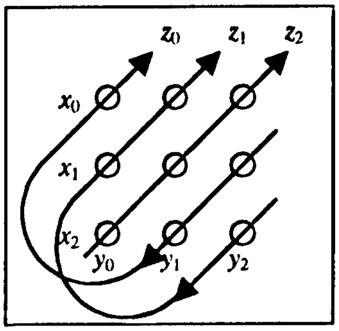


Figure 1.

An association between two items **x** and **y** can be represented by the convolution of the two items: **x**⊗**y**. The inverse vector of **x** can be used to reconstruct **y** from **x**⊗**y**: $\mathbf{x}^{-1}$⊗(**x**⊗**y**)=**y**. However, except under certain restrictive conditions, the inverse is numerically unstable and is not always the best choice for decoding. For vectors which have randomly chosen elements independently distributed as N(0,1/*n*) (the normal distribution with mean 0 and variance 1/*n*) there is an approximate inverse with attractive properties. The approximate inverse of **x** is denoted by $\mathbf{x}^{T}$. It is a simple rearrangement of the elements of **x**: $x^{T}_{i}=x_{-i}$, where subscripts are modulo-*n*. The approximate inverse is simple to compute and is numerically stable. Reconstruction using the approximation inverse is noisy. The convolution product $\mathbf{x}^{T}$⊗**x**⊗**y** can be written as **y**+η, where the η can be considered as zero-mean noise whose magnitude (variance) decreases with increasing vector dimension.

Multiple associations can be represented by the sum of the individual associations. For example, suppose **x**, **y**, **v**, and **w** are all randomly chosen vectors with elements independently distributed as N(0,1/*n*). The association of **x** with **y** and **v** with **w** can be represented by **z**=**x**⊗**y**+**v**⊗**w**. To find what is associated with **x** we convolve **z** with $\mathbf{x}^{T}$. The result can be expressed as $\mathbf{x}^{T}$⊗**x**⊗**y**+$\mathbf{x}^{T}$⊗**v**⊗**w**. The first term is approximately equal to **y** and the second term can be regarded as noise - it will not be highly correlated with any of **x**, **y**, **v**, or **w**. The sum of the two terms will be recognizable as a distorted version of **y**.

3.3 Similarity preservation and randomization

Convolution preserves both similarity and lack of similarity in a multiplicative fashion: the similarity of two role-filler binding patterns is approximately equal to the product of the similarities of the respective role and filler patterns (provided that the role patterns are not similar to the filler patterns.) Thus, if two bindings have the same role, their similarity will be equal to that of the fillers. Conversely, if two roles have no similarity, bindings involving them will have similarity regardless of the fillers. Furthermore, convolution is randomizing in that role-filler binding patterns are not similar to either the role or filler patterns.

3.4 HRRs for relational structure

Consider representing a nested proposition such as "Spot bit Jane, causing Jane to flee from Spot" in a vector pattern. We would like this pattern to faithfully record structure and also to be suitable for detecting at least superficial similarity by computing dot-products.

The structure of a proposition can be represented by superimposing patterns representing the predicate name and the role-filler bindings. This provides a structural skeleton that faithfully records structure.

The skeleton HRR for the proposition "Spot bit Jane" is constructed as follows:

$$\mathbf{K}_{\mathbf{P\text{-}bite}}=\mathbf{bite}+\mathbf{bite}_{agt}\otimes\mathbf{spot}+\mathbf{bite}_{obj}\otimes\mathbf{jane}$$

The pattern **bite** represents the predicate label, $\mathbf{bite}_{agt}$ and $\mathbf{bite}_{obj}$ its roles, and **spot** and **jane** the entities "Spot" and "Jane". If we have the pattern $\mathbf{K}_{\mathbf{P\text{-}bite}}$ and know the role patterns, then we can reconstruct the filler patterns by convolving $\mathbf{K}_{\mathbf{P\text{-}bite}}$ with the approximate inverses of the role patterns. For example, $\mathbf{bite}_{agt}^{T}\otimes\mathbf{K}_{\mathbf{P\text{-}bite}}$ gives a noisy version of **spot** which, if necessary, can be cleaned up using an auto-associative item memory. The pattern **bite** is made a component of $\mathbf{K}_{\mathbf{P\text{-}bite}}$ in order to identify it as a *bite* proposition and thus allow a system to deduce that the appropriate role patterns for decoding are $\mathbf{bite}_{agt}$ and $\mathbf{bite}_{obj}$.

The skeleton HRR pattern for the proposition "Spot bit Jane" is an *n*-dimensional pattern just like the patterns **spot**, **bite**, etc. Thus, it is easily used as a filler in a higher-order proposition. For example, the skeleton HRR $\mathbf{K}_{\mathbf{P}}$ representing "Spot bit Jane, which caused Jane to flee from Spot" is constructed as follows:

$$\mathbf{K}_{\mathbf{P\text{-}flee}}=\mathbf{flee}+\mathbf{flee}_{agt}\otimes\mathbf{jane}+\mathbf{flee}_{from}\otimes\mathbf{spot}$$
$$\mathbf{K}_{\mathbf{P}}=\mathbf{cause}+\mathbf{cause}_{antc}\otimes\mathbf{K}_{\mathbf{P\text{-}bite}}+\mathbf{cause}_{cnsq}\otimes\mathbf{K}_{\mathbf{P\text{-}flee}}$$

The other goal for a vector representation was that patterns should reflect superficial similarity, i.e., two patterns should be similar if the structures they represent merely involve similar fillers or predicates. The presence of predicate labels in HRRs ensures that patterns for the same predicate are similar. However, skeleton HRRs do not behave as desired with respect to the presence of similar fillers: the randomizing properties of convolution mean that $\mathbf{role}_1\otimes\mathbf{filler}_1$ is only similar to $\mathbf{role}_2\otimes\mathbf{filler}_2$ to the extent that $\mathbf{role}_1$ is similar to $\mathbf{role}_2$ *and* $\mathbf{filler}_1$ is similar to $\mathbf{filler}_2$. HRRs are easily made to reflect superficial similarity by superimposing the filler patterns together with the structural skeleton HRR. Thus, the fleshed-out HRR for "Spot bit Jane" is as follows:

$$\begin{aligned}\mathbf{P}_{\mathbf{bite}} = \;&\mathbf{bite}+\mathbf{spot}+\mathbf{jan}+\\ &+\mathbf{bite}_{agt}\otimes\mathbf{spot}+\mathbf{bite}_{obj}\otimes\mathbf{jane}\end{aligned}$$

Adding in the fillers makes decoding more noisy, but does not prevent successful decoding. For higher level propositions, the same idea of adding in fillers can be applied recursively. For example, the HRR for "Spot bit Jane, causing Jane to flee from Spot" is constructed as follows:

$$\begin{aligned}\mathbf{P}_{\mathbf{flee}} = \;&\mathbf{flee}+\mathbf{spot}+\mathbf{jane}+\\ &+\mathbf{flee}_{agt}\otimes\mathbf{jane}+\mathbf{flee}_{from}\otimes\mathbf{spot}\\ \mathbf{P} = \;&\mathbf{cause}+\mathbf{P}_{\mathbf{bite}}+\mathbf{P}_{\mathbf{flee}}+\\ &+\mathbf{cause}_{antc}\otimes\mathbf{P}_{\mathbf{bite}}+\mathbf{cause}_{cnsq}\otimes\mathbf{P}_{\mathbf{flee}}\end{aligned}$$

HRRs constructed like this will be similar if they merely involve similar entities or predicates. Because of the similarity preserving properties of convolution, they will be even more similar if the entities are involved in similar roles.

3.5 The need for a "clean-up" memory

Convolution encodings are remarkably compact: a number of associations between *n*-dimensional patterns packed into one *n*-dimensional pattern. The price we pay for this compactness is noise in decoded vectors. Consequently, if we want a convolution-based associative-memory model to provide accurate reconstructions of decoded patterns, it must be equipped with an additional error-correcting auto-associative item memory. This can clean up the noisy patterns retrieved from the convolution encodings. This clean-up memory must store all the items that the system can produce. When given a noisy version of one of those items it must either output the closest item or indicate that the input is not close to any of the stored items. Note that only a few associations are stored as convolution encodings in a single pattern, whereas many patterns are stored in the clean-up memory.

3.6 Normalization

The final point to consider when constructing HRRs is maintaining the overall strength of patterns and the statistical distribution of their elements. The easiest way to do this is to normalize all patterns to have a Euclidean length of one. Here, the normalized version of the vector **x** is denoted by $\langle \mathbf{x} \rangle$ and is defined as follows:

$$\langle \mathbf{x} \rangle = \mathbf{x} / \sqrt{\sum_{i=0}^{n-1} x_i^2}$$

4. ESTIMATING SIMILARITY

The six "dog bites human" episodes provide a simple demonstration that HRR scores can reflect similarity of structural arrangements, as well as similarity of surface features. It also demonstrates that a model based just on HRR similarity scores can neatly explain the pattern of human retrieval: LS > CM ≥ SF > AN ≥ FOR.

The HRRs for the probe (**P**) and the literally similar episode ($\mathbf{E}_{LS}$) are constructed as follows:

$$\begin{aligned}
\mathbf{P}_{bite} &= \langle \mathbf{bite} + \langle \mathbf{spot} + \mathbf{jane} \rangle \\
&\quad + \mathbf{bite}_{agt} \circledast \mathbf{spot} + \mathbf{bite}_{obj} \circledast \mathbf{jane} \rangle \\
\mathbf{P}_{flee} &= \langle \mathbf{flee} + \langle \mathbf{spot} + \mathbf{jane} \rangle \\
&\quad + \mathbf{flee}_{agt} \circledast \mathbf{jane} + \mathbf{flee}_{from} \circledast \mathbf{spot} \rangle \\
\mathbf{P} &= \langle \mathbf{cause} + \langle \mathbf{P}_{bite} + \mathbf{P}_{flee} \rangle \\
&\quad + \mathbf{cause}_{antc} \circledast \mathbf{P}_{bite} + \mathbf{cause}_{cnsq} \circledast \mathbf{P}_{flee} \rangle \\
\mathbf{E}_{LS\text{-}bite} &= \langle \mathbf{bite} + \langle \mathbf{spot} + \mathbf{jane} \rangle \\
&\quad + \mathbf{bite}_{agt} \circledast \mathbf{spot} + \mathbf{bite}_{obj} \circledast \mathbf{jane} \rangle \\
\mathbf{E}_{LS\text{-}flee} &= \langle \mathbf{flee} + \langle \mathbf{spot} + \mathbf{jane} \rangle \\
&\quad + \mathbf{flee}_{agt} \circledast \mathbf{jane} + \mathbf{flee}_{from} \circledast \mathbf{spot} \rangle \\
\mathbf{E}_{LS} &= \langle \mathbf{cause} + \langle \mathbf{P}_{bite} + \mathbf{P}_{flee} \rangle \\
&\quad + \mathbf{cause}_{antc} \circledast \mathbf{P}_{bite} + \mathbf{cause}_{cnsq} \circledast \mathbf{P}_{flee} \rangle
\end{aligned}$$

The HRRs for the other episodes are built in an analogous fashion. The patterns for members of the same species (types) are designed to be similar. The complete set of base vectors and tokens used in this experiment is shown in Table 1. All base and identity (**id**) vectors were randomly chosen with elements independently distributed as N(0,1/*n*).

Average HRR similarity scores are shown in Table 2. These are from 100 runs with different random base and identity vectors, and a vector dimension of 2048. The directions of differences between average similarity scores were reliable - the standard deviation of the scores ranged between 0.016 and 0.026.

Base vectors		Token vectors
person	**bite**	**jane** = $\langle$**person** + $\mathbf{id}_{jane}\rangle$
dog	**flee**	**john** = $\langle$**person** + $\mathbf{id}_{john}\rangle$
cat	**cause**	**fred** = $\langle$**person** + $\mathbf{id}_{fred}\rangle$
mouse		**spot** = $\langle$**dog** + $\mathbf{id}_{spot}\rangle$
		fido = $\langle$**dog** + $\mathbf{id}_{fido}\rangle$
$\mathbf{bite}_{agt}$	$\mathbf{bite}_{obj}$	**rover** = $\langle$**dog** + $\mathbf{id}_{rover}\rangle$
$\mathbf{flee}_{agt}$	$\mathbf{flee}_{obj}$	**felix** = $\langle$**cat** + $\mathbf{id}_{felix}\rangle$
$\mathbf{cause}_{antc}$	$\mathbf{cause}_{cnsq}$	**mort** = $\langle$**mouse** + $\mathbf{id}_{mort}\rangle$

Table 1.

For comparison, MAC-style similarity scores are also shown. These are modeled after the MAC stage of Forbus *et al*'s (1994) MAC/FAC model. They are based on the dot product of normalized content vectors over the following features: *person*, *dog*, *mouse*, *cat*, *cause*, *bite*, and *flee*. For example, the content vector for the probe is (1,1,0,0,1,1,1)/√5.

The pair of episodes $\mathbf{E}_{LS}$ and $\mathbf{E}_{SF}$ each have the same surface commonalities (object features and predicate names) with the probe. The difference between them is that $\mathbf{E}_{LS}$ is structurally isomorphic to the probe, while $\mathbf{E}_{SF}$ is not. Because there is no structural information beyond predicates names encoded in content vectors, $\mathbf{E}_{LS}$ and $\mathbf{E}_{SF}$ have the same content-vector similarity to the probe. On the other hand, the HRR similarity scores indicates that $\mathbf{E}_{LS}$ is more similar to the probe than $\mathbf{E}_{SF}$.

When episodes do not share object attributes with the probe, HRR scores are low and do not always reflect structural match. Although in Table 2 the HRR score for $\mathbf{E}_{AN}$ is higher than for $\mathbf{E}_{FOR}$ (due to the "bite" and "flee" propositions filling the same roles in $\mathbf{E}_{AN}$ as in the probe), this difference is not reliable. It is possible to construct other FOR examples that have a higher score than AN examples (Plate, 1994).

$\mathbf{E}_{CM}$ is a cross-mapped analogy. It has the same structure and types of objects as the probe, but unlike $\mathbf{E}_{LS}$ and the probe, the similar objects do not map to each other (the dog maps to the person, and the person maps to the dog). Since HRR similarity scores are sensitive to having similar objects fill similar roles, $\mathbf{E}_{CM}$ has a lower HRR similarity to the probe than $\mathbf{E}_{LS}$. In contrast, the content-vector similarities of $\mathbf{E}_{CM}$ and $\mathbf{E}_{LS}$ to the probe are the same.

4.1 Why HRR dot-products reflect structural similarity

HRR dot-products reflect structural similarity because of the presence of components representing combinatorial features, such as $\mathbf{bite}_{agt}\otimes\mathbf{spot}$, $\mathbf{cause}_{antc}\otimes\mathbf{bite}$, and $\mathbf{cause}_{antc}\otimes\mathbf{bite}_{agt}\otimes\mathbf{spot}$.

All of these higher-order features derive from role-filler bindings. Consequently, the HRRs described here reflect differences in structural similarity when there are differences in whether similar objects fill similar roles. Hence the large difference between $\mathbf{E}_{CM}$ and $\mathbf{E}_{LS}$ in their HRR similarity scores with the probe. Although they have the same objects, and isomorphic structure, $\mathbf{E}_{CM}$ does not have similar objects filling the same roles as in the probe. Thus, $\mathbf{E}_{CM}$ has combinatorial features like $\mathbf{bite}_{agt}\otimes\mathbf{person}$, which are not at all similar to those like $\mathbf{bite}_{agt}\otimes\mathbf{dog}$.

This pattern of sensitivity to structural similarity, in which structural similarity is only detected when similar objects fill similar roles, is very similar to the pattern observed by Ross (1989) in experiments with people. Ross found that shared structure enhanced retrieval in the presence of similar objects, provided that corresponding objects were similar, and that cross-mapping inhibited retrieval.

5. INTERPRETATIONS OF AN ANALOGY

Retrieval of analogies is only the first step in many analogy processing tasks. After retrieving a potentially analogous episode we may want to decode the structure in order to evaluate more accurately the degree of structural consistency, or to use the episode for analogical reasoning. The structure of a HRR could be decoded using the techniques described in Section 2, and then used in a symbolic processor like SME or in some other connectionist architecture. However, some apparently more symbolic tasks, like finding corresponding entities, and thus deriving an interpretation of an analogy, can be computed with vector operations directly on HRRs.

Consider the probe **P** "Spot bit Jane, causing Jane to flee from Spot", and $\mathbf{E}_{LS}$ "Fido bit John, causing John to flee from Fido." The entity corresponding to Jane (which is John) can be found in two steps:

		Commonalities with probe			Similarity scores	
P	Spot bit Jane, causing Jane to flee from Spot.	Object attributes	First-order relation names	Higher-order structure		
	Episodes in long-term memory:				HRR	MAC
$\mathbf{E}_{LS}$	Fido bit John, causing John to flee from Fido.	✓	✓	✓	0.71	1.0
$\mathbf{E}_{SF}$	John fled from Fido, causing Fido to bite John	✓	✓	×	0.47	1.0
$\mathbf{E}_{CM}$	Fred bit Rover, causing Rover to flee from Fred.	✓	✓	✓	0.47	1.0
$\mathbf{E}_{AN}$	Mort bit Felix, causing Felix to flee from Mort.	×	✓	✓	0.42	0.6
$\mathbf{E}_{FOR}$	Mort fled from Felix, causing Felix to bite Mort.	×	✓	×	0.30	0.6

Table 2.

1. Extract the roles Jane fills in the probe with the operation:

$\langle \mathbf{P} \otimes \mathbf{jane}^{T} \rangle = \mathbf{jane\text{-}roles}$

This pattern is a blend of various roles and other noise patterns. The following are the positive dot-products of the **jane-roles** pattern with other role patterns:

$\mathbf{jane\text{-}roles} \times \mathbf{cause}_{antc} = 0.20$

$\mathbf{jane\text{-}roles} \times \mathbf{cause}_{cnsq} = 0.18$

$\mathbf{jane\text{-}roles} \times \mathbf{flee}_{agt} = 0.13$

$\mathbf{jane\text{-}roles} \times \mathbf{bite}_{obj} = 0.12$

2. Use **jane-roles** to extract the fillers from $\mathbf{E}_{LS}$ and compare with the entities in $\mathbf{E}_{LS}$:

$\langle \mathbf{E}_{LS} \otimes \mathbf{jane\text{-}roles}^{T} \rangle \times \mathbf{john} = 0.38$

$\langle \mathbf{E}_{LS} \otimes \mathbf{jane\text{-}roles}^{T} \rangle \times \mathbf{fido} = 0.05$

The most similar pattern is **john**, which is in fact the entity in $\mathbf{E}_{LS}$ corresponding to Jane.

$\langle \mathbf{E}_{LS} \otimes \mathbf{jane\text{-}roles}^{T} \rangle$	**john** **fido**	0.38 0.07	✓
$\langle \mathbf{E}_{CM} \otimes \mathbf{jane\text{-}roles}^{T} \rangle$	**fred** **rover**	0.25 0.17	×
$\langle \mathbf{E}_{AN} \otimes \mathbf{jane\text{-}roles}^{T} \rangle$	**felix** **mort**	0.16 0.09	✓
$\langle \mathbf{E}_{SF} \otimes \mathbf{jane\text{-}roles}^{T} \rangle$	**john** **fido**	0.23 0.07	?
$\langle \mathbf{E}_{FOR} \otimes \mathbf{jane\text{-}roles}^{T} \rangle$	**mort** **felix**	0.11 0.06	?

Table 3.

Table 3 shows the extraction of the entities corresponding to Jane in the various episodes. Correct extractions are checkmarked, and cases where there is no clear corresponding object have a question mark.

The correct answer is obtained in $\mathbf{E}_{LS}$, where corresponding objects are similar, and in $\mathbf{E}_{AN}$, where there is no object similarity. This extraction process has a bias towards choosing similar entities as the corresponding ones, which leads to a reasonable answer for $\mathbf{E}_{SF}$ and an incorrect answer $\mathbf{E}_{CM}$. There are no correct answers for $\mathbf{E}_{SF}$ and $\mathbf{E}_{FOR}$, because there are no consistent mapping between $\mathbf{P}$ and those episodes. However, because of the bias for mapping similar items, Fred is strongly indicated to be the entity in $\mathbf{E}_{SF}$ corresponding to Jane. The only wrong answer is given for the cross-mapped analogy $\mathbf{E}_{CM}$ where again the more similar object is indicated to be the corresponding one. Again, the effect of cross-mapping is similar to that observed by Ross (1989) in people: cross-mapping causes less accurate mapping performance.

Closer examination of the extraction process reveals both the reason for this bias and several ways of eliminating it, if that should be desired. Consider patterns containing just two of the components from $\mathbf{P}$ and $\mathbf{E}_{CM}$:

$$\mathbf{P}' = \mathbf{cause} + \mathbf{bite}_{obj} \otimes \mathbf{jane}$$
$$\mathbf{E}'_{CM} = \mathbf{cause} + \mathbf{bite}_{obj} \otimes \mathbf{rover}$$

The roles of Jane in $\mathbf{P}'$ are computed as:

$$\begin{aligned} \mathbf{jane\text{-}roles}' &= \mathbf{P}' \otimes \mathbf{jane}^{T} \\ &\approx \mathbf{cause} \otimes \mathbf{jane}^{T} + \mathbf{bite}_{obj} \end{aligned}$$

The role pattern $\mathbf{bite}_{obj}$ (and other role patterns like $\mathbf{flee}_{agt}$ and $\mathbf{cause}_{antc}$ in the full version of $\mathbf{P} \otimes \mathbf{jane}^{T}$) are what are wanted here. The other patterns like $\mathbf{cause} \otimes \mathbf{jane}^{T}$, which are not roles at all, are the source of the same-type bias in finding the corresponding object. When $\mathbf{jane\text{-}roles}^{T}$ is used to extract the fillers from $\mathbf{E}_{CM}$, we get the following:

$$\begin{aligned} \mathbf{corresp} &= \mathbf{jane\text{-}roles}^{T} \otimes \mathbf{E}'_{CM} \\ &\approx (\mathbf{cause} \otimes \mathbf{jane}^{T} + \mathbf{bite}_{obj})^{T} \\ &\otimes (\mathbf{cause} + \mathbf{bite}_{obj} \otimes \mathbf{rover}) \\ &= \underline{\mathbf{jane}} + \mathbf{cause} \otimes \mathbf{jane}^{T} \otimes \mathbf{bite}_{obj} \otimes \mathbf{rover} \\ &\quad + \mathbf{bite}_{obj}^{T} \otimes \mathbf{cause} + \underline{\mathbf{rover}} \end{aligned}$$

This includes the pattern **rover** as desired, but also includes the pattern **jane** (from $(\mathbf{cause} \otimes \mathbf{jane}^{T})^{T} \otimes \mathbf{cause}$). Although **corresp'** only contains one term like this, there is a **jane** component in **corresp** for every pattern which is shared by $\mathbf{P}$ and $\mathbf{E}_{CM}$. This adds up to a very strong component of **jane** in **corresp**. When **corresp** is compared to the fillers of $\mathbf{E}_{CM}$, **corresp** is more similar to **fred** than **rover**, due to the strong **jane** component in **corresp**.

One way of eliminating this similar-type bias is to perform a linear, multi-way, role-clean-up on **jane-roles**. This should pass all positive role components and suppress negative role and non-role components like $\mathbf{cause} \otimes \mathbf{jane}^{T}$. Thus, the clean version of **jane-roles** is as follows:

$$\mathbf{clean\text{-}jane\text{-}roles} = 0.20\times\mathbf{cause}_{\mathrm{antc}} + 0.18\times\mathbf{cause}_{\mathrm{cnsq}} + 0.13\times\mathbf{flee}_{\mathrm{agt}} + 0.12\times\mathbf{bite}_{\mathrm{obj}}$$

$\langle \mathbf{E}_{\mathrm{LS}} \otimes \mathbf{cleaned\text{-}jane\text{-}roles}^{\mathrm{T}} \rangle$	**john**	0.27	✓
	fido	0.20	
$\langle \mathbf{E}_{\mathrm{CM}} \otimes \mathbf{cleaned\text{-}jane\text{-}roles}^{\mathrm{T}} \rangle$	**fred**	0.20	
	rover	0.29	✓
$\langle \mathbf{E}_{\mathrm{AN}} \otimes \mathbf{cleaned\text{-}jane\text{-}roles}^{\mathrm{T}} \rangle$	**felix**	0.25	✓
	mort	0.20	
$\langle \mathbf{E}_{\mathrm{SF}} \otimes \mathbf{cleaned\text{-}jane\text{-}roles}^{\mathrm{T}} \rangle$	**john**	0.25	?
	fido	0.17	
$\langle \mathbf{E}_{\mathrm{FOR}} \otimes \mathbf{cleaned\text{-}jane\text{-}roles}^{\mathrm{T}} \rangle$	**mort**	0.26	?
	felix	0.19	

Table 4.

The corresponding objects extracted using role clean-up are shown in Table 3. This slightly slower process gives correct answers for the episodes in which there is a consistent mapping.

The other way of avoiding the similar-type bias is to use a different binding operation, in which the algebraic properties of encoding and decoding do not result in terms like $(\mathbf{cause}\otimes\mathbf{jane}^{\mathrm{T}})^{\mathrm{T}}\otimes\mathbf{cause}$ equating to **jane**. Possible suitable alternative binding operations are discussed in Plate (1994).

There are two more limitations with these fast techniques for deriving interpretations. One is that each corresponding pair in a mapping is extracted independently. This matters when there is more than one consistent mapping. For example, if we have two possible consistent mappings $\{X \leftrightarrow A,\ Y \leftrightarrow B\}$ and $\{X \leftrightarrow B,\ Y \leftrightarrow A\}$, then the choice of mapping for X should constrain the choice for Y, but this will not be the case with the above techniques. To overcome this problem requires some other mechanism for checking that a mapping is one-to-one. The other problem is that these techniques fail when two different objects have the same set of roles - in such a case ambiguous results can be produced.

6. CHUNKING & MEMORY ORGANIZATION

HRRs provide a natural method for chunking. In fact, a model based on HRRs must use chunking if it is to store structures of unlimited size. Chunking involves storing sub-structures in the item memory, and using them when decoding components of complex structures. For example, to decode the agent of the cause antecedent of **P** we first extract the cause antecedent pattern. This gives a noisy version of $\mathbf{P}_{\mathrm{bite}}$, which can be cleaned up by accessing item memory and retrieving the closest match. Now we have an accurate version of $\mathbf{P}_{\mathrm{bite}}$ from which we can extract the filler of the agent role.

To use chunks there must be a way of referring, or pointing to the chunks. In content-addressable memory in general, "pointers" to sub-chunks cannot be addresses, but must somehow hint at the contents of the sub-chunk. In HRRs, a decoded filler or sub-chunk, which is derived from a chunk by decoding with a role pattern, functions as an associative "pointer" to a pattern in item memory. These associative pointers are different from conventional pointers in that their form conveys information about their referent, information that is noisy but immediately available without the need to access memory. The advantage of having pointers that encode information about their referents is that some operations can be performed without following the pointer. This can save much time. For example, we can decode nested fillers quickly if very noisy results are acceptable, or we can get an estimate of the similarity of two structures without decoding them.

6.1 Overall memory organization

In a system that uses HRRs there must be two levels of memory organization. One level encodes the structure in and among chunks. The other level stores large numbers of chunks (the large-scale clean-up memory).

Convolution encoding is most suited for encoding structure in and among small chunks in memory. Because of its memory capacity characteristics and noise in retrieval, convolution does not provide a suitable associative memory technique for the clean-up memory, which must store all the chunks. For this purpose we require some sort of large-scale error-correcting auto-associative memory. This large-scale memory should have the following properties:

> **auto-associative & error-correcting ability**: when given a pattern, it should return accurately the closest one(s) stored, and
>
> **high capacity**: the number of patterns which can be stored should be exponential in the size of the patterns.

There are several ways the clean-up memory could be implemented, e.g., Kanerva's (1988) sparse distributed memory, and Baum *et al*'s (1988) various content addressable schemes.

7. DISCUSSION

This paper has described a scheme for encoding structure in vector representations based on circular convolution. Other approaches, such as Smolensky's (1990) tensor products, Pollack-'s (1990) RAAMs, Kanerva's (1996) binary spattercodes, have much in common - see Plate (1997) for a discussion - and could also be used in models of analogy processing.

The origin of patterns representing types such as 'dog', 'cat' and 'human' must be addressed at some stage. One possible automatic technique for learning such patterns is Latent Semantic Analysis (LSA), which learns high-dimensional vector patterns for words from large quantities of text (Landauer, Laham, and Foltz 1998). These patterns reflect human similarity judgements and could easily be used with HRRs.

The existence of a fast technique for computing good guesses at object correspondences suggests a new model for analogical mapping. Mapping could be done by "guessing" correspondences while stepping through the components of two structures and verifying that the proposed correspondences are consistent. This would require three mechanisms, one for traversing structures, another for guessing correspondences, and the last for storing correspondences and checking their consistency. All can be implemented with operations on vector representations. Such a model differs from ACME and SME in that it puts complexity at a different level. The top level involves simple sequential computation (traversing a structure and checking for mapping inconsistencies) rather than complex structural matching or construction of special networks, while the bottom level involves information-rich vector processing to measure similarities and estimate correspondences.

8. CONCLUSION

Holographic Reduced Representations provide a useful vector representation for analog retrieval and processing tasks. They provide chunking, which will be essential in vector-based model that stores large structures. They also support fast operations for computing similarity and object correspondences. These fast operations appear to have the right amount of power for modeling human abilities: their strengths and weaknesses follow a similar pattern to human performance on various analogy tasks. In particular, HRRs provide a simple, single-stage model of human performance on analog retrieval: HRR dot-products are sensitive to superficial similarity, and also to structural similarity in situations where corresponding roles have similar fillers, which is the same pattern of performance as demonstrated by human subjects on analog retrieval tasks.

9. REFERENCES

Baum, E. B., J. Moody, and F. Wilczek (1988). Internal representations for associative memory. *Biological Cybernetics* 59, 217-228.

Fodor, J. A. and Z. W. Pylyshyn (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* 28, 3-71.

Forbus, K. D., D. Gentner, and K. Law (1994). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science* 19, 141-205.

Gentner, D. and A. B. Markman (1993). Analogy - Watershed or Waterloo? Structural alignment and the development of connectionist models of analogy. In C. L. Giles, S. J. Hanson, and J. D. Cowan (Eds.), *Advances in Neural Information*

Processing Systems 5 (NIPS'92), CA, 855-862. Morgan Kaufmann.

Gentner, D., M. J. Rattermann, and K. D. Forbus (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology* 25(4), 524-575.

Hinton, G. E., J. L. McClelland, and D. E. Rumelhart (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, and the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, Volume 1, 77-109. MIT Press.

Kanerva, P. (1988). *Sparse Distributed Memory.* MIT Press.

Kanerva, P. (1996). Binary spatter-coding of ordered k-tuples. In C. von der Malsburg, W. von Seelen, J. Vorbruggen, and B. Sendhoff (Eds.), Artificial Neural Networks-ICANN Proceedings, Volume 1112 of *Lecture Notes in Computer Science, Berlin*, 869-873. Springer.

Landauer, T. K., D. Laham and P. W. Foltz (1998). Learning Human-like Knowledge with Singular Value Decomposition: A Progress Report. In *Neural Information Processing Systems* (NIPS'97). MIT Press.

Plate, T. A. (1994). Distributed Representations and Nested Compositional Structure. Ph.D. thesis, Dept. of Computer Science, University of Toronto. Available at http://www.mcs.vuw.ac.nz/~tap.

Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks* 6(3), 623-641.

Pollack, J. B. (1990). Recursive distributed representations. *Artificial Intelligence* 46(1-2), 77-105.

Ratcliff, R. and G. McKoon (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology* 21, 138-155.

Ross, B. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology : Learning, Memory, and Cognition* 15(2), 456-468.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46(1-2), 159-216.

Thagard, P., K. J. Holyoak, G. Nelson, and D. Gochfeld (1990). Analog Retrieval by Constraint Satisfaction. *Artificial Intelligence* 46, 259-310.

Wharton, C. M., K. J. Holyoak, P. E. Downing, T. E. Lange, T. D. Wickens, and E. R. Melz (1994). Below the surface: Analogical similarity and retrieval competition in reminding. *Cognitive Psychology* 26, 64-101.

DUAL ROLE OF ANALOGY IN THE DESIGN OF A COGNITIVE COMPUTER

Pentti Kanerva
RWCP[1] Theoretical Foundation SICS[2] Laboratory
SICS, Box 1263, SE-164 29 Kista, Sweden
E-mail: kanerva@sics.se

ABSTRACT

This paper is about the computer analogy of the brain and how it can both help and hinder our understanding of the human mind. It is based on the assumptions that the mind can be understood in terms of the working of the brain, and that the brains function is to process information: that it is some kind of a computer, as contrasted for example with the heart which is a pump. It is a computer whose design we do not understand but try to, by analogy; that is, by making a model – a "cognitive computer" – based on our understanding of computers, brains, and the working of the mind.

Human intelligence and language are fundamentally analogical and figurative whereas lower forms of intelligence and conventional computers treat meaning literally. Therefore, the challenge in designing a cognitive computer is to find the kinds of information representation and operations that make figurative meaning come out naturally. The paper discusses holistic representation, which is unconventional and looks promising and worthy of investigation – it easily encodes recursive (list) structure, for example – and points out a danger in taking too literally cognitive models that have been developed on conventional computers, such as the following of rules.

INTRODUCTION

The human mind is unlike any computer or program we know. It is not literal, and when meaning is taken literally, the result can be funny or total nonsense. Thats the humor of puns. This must mean that the human mind, although capable of being literal, is fundamentally figurative or symbolical or analogical. How else could we judge a literal interpretation as being at once both accurate and wrong?

The growth of the human mind – our grasp of things – is largely due to analogical perceiving and thinking. Some things are meaningful to us at birth or without learning: they are mostly things necessary for survival. The rest we learn through experience. Some learning is associative, as when we learn cause and effect. This kind of learning is basic to all animals.

To follow an example, or imitate, is a more advanced form of learning and is common at least in mammals and birds. It involves a basic form of analogy. The learner identifies with a role model – perceives one as the other, makes an analogical connection or mapping between oneself and the other.

Full-fledged analogy is central to human intelligence. We relate the unfamiliar to the familiar, and we see the new in terms of the old. This is most evident in language, which is thoroughly metaphorical. New and unfamiliar things are expressed and explained in familiar terms that are understood not literally but figuratively. It is possible that full-fledged analogy and human language need each other and that our faculties for them have coevolved.

Analogy is such an integral part of us that we hardly notice it nor pay it its proper dues. That is, until we try to program a computer to act like a human. AI has puzzled over the programming of humanlike behavior for three decades. At first it was thought that programming computers to understand language, to translate, and the like, were just around the corner, waiting only for computers to get large and fast enough. Now they are large and fast, many things have been tried and much has been learned, but the puzzle remains and we have no clear idea of how to solve it.

---

[1] Real World Computing Partnership. Support of this research by Japans Ministry of International Trade and Industry through the RWCP is gratefully acknowledged.

[2] Swedish Institute of Computer Science.

This paper is a personal view of the lessons this holds for us. The theme is that we must rethink computing, put figurative meaning and analogy at its center, and find computing mechanisms that make it come out naturally. This can be construed as designing a new kind of computer, a "cognitive computer," that is a better model of the brain than present-day computers are. I will also try to verbalize things that students of connectionist architectures take for granted but that might puzzle others, the main idea being that implementation matters when we try to understand how the mind works.

THE COMPUTER AS A BRAIN AND THE BRAIN AS A COMPUTER

Equating computers with brains is an example of analogical thinking. Early computers were dubbed electronic brains, computers have memory, and we even say that a program knows, wants, or believes so and so. Such anthropomorphizing seems natural to us and it serves a purpose. It brings a technological mystery within the realm of the familiar, since we already have an idea of what the brain does even if we dont know just how it does it.

We also talk of the brain as a computer. Its appeal is in that whereas the mechanisms of the brain are hidden, those of the computer are available to us, and through them we could possibly understand the brains mechanisms. The principle is sound and is the thesis behind Turings imitation game: If we can build a machine that behaves in the same way as a natural system does, we have understood the natural system.

Analogies not only help our thinking but they also channel and limit it. The computer analogy of the brain or of the mind has certainly done so, as modeling in cognitive science and AI has been dominated by programs written for the computer, while philosophical and qualitative treatment of issues is looked upon with suspicion.

Many things are modeled successfully on computers, such as weather, traffic flow, strength of materials and structures, industrial processes, and so forth. However, there are special pitfalls when the thing being modeled "the brain" is itself some kind of a computer: the danger is that our models begin to look like the computers they run on or the programming languages they are written in. For example, we talk of human short-term or working memory and think of the computers active registers, or we talk of human long-term memory and think of the computers permanent storage (RAM or disk), or we talk of the grammar of a language and think of a tree-structure or a set of rewriting rules programmed in Lisp. Of course these are analogical counterparts, but there is a danger of taking them too literally. Human memory works very differently from computer memory, and the brain is not a Lisp machine nor the mind a logic program. Some analogical comparisons have not been at all useful in understanding the working of the mind: for example, equating the brain with the computers hardware and the mind with its software. Finally, there is a worse danger of failing to notice what is missing in our models of the mind because it is missing or invisible in computers. To safeguard against it, we must treat the subject qualitatively: Our models may behave as advertised, but is that how people behave; for example, how they use language?

ARTIFICIAL NEURAL NETS AS BIOLOGICALLY MOTIVATED MODELS OF COMPUTING

The computers and brains architectures are very different. Perhaps the differences account for the difficulty in programming computers to be more lifelike and less literal-minded. This has motivated the study of alternative computing architectures called (artificial) neural nets (NN), or parallel distributed processing (PDP), or connectionist architectures. The hope is that an architecture more similar to the brains should produce behavior more similar to the brains, which is a valid analogical argument. Unfortunately it does not tell us what in the architecture matters and what is incidental, and unfortunately our neural nets are not significantly more figurative than traditional computers.

Neural-net research has made a valuable contribution by focusing our attention on representation. Computer theoreticians and engineers know, for example, that the representation of numbers has a major effect on circuit design. A representation that works well for addition works reasonably well also for multiplication, whereas a representation that allows very fast multiplication is useless for addition. Thus a representation is a compromise that favors some operations and hinders others.

Information in computers is stored locally, that is, in records with fields. Local representation – one unit per concept – is common also in neural nets. The alternative is to distribute information from many sources over shared units. It is more brainlike, at least superficially, and it has been studied and used with neural nets for a long time. I take distributed representation to be fundamental to the brains operation and believe that a cognitive computer should be based on it, and that therefore we should find out all we can about the encoding of information into, and operating with, distributed representations.

Neural-net research has shown that these representations are robust and support some forms of generalization: representations (patterns) that are similar on the surface – close according to some metric – are treated similarly, for example as belonging in the same or similar classes. The representations are also suitable for learning from examples. The learning takes place by statistical averaging or clustering of representations (self-organizing). It is not very creative but it can be subtle and lifelike, which makes it cognitively interesting. It can produce behavior that looks like rule-following although the system has no explicit rules, as was demonstrated with the learning of the past tense of English verbs by Rumelhart and McClelland (1986). This is a significant discovery, in that it demonstrates a principle that probably governs the working of the brain in general and should govern the working of a cognitive computer. What we see and describe as rule-following is an emergent phenomenon that reflects an underlying mechanism. However, the rules do not produce the behavior even if they may accurately describe it.

DESCRIPTION VS. EXPLANATION

The distinction between description and explanation of behavior is so central that I will highlight it with an example. Consider heredity. Long before the genetic bases of heredity were known, people knew about dominant and recessive traits and had figured out the basic laws of inheritance. For example, a plant species may come in three varieties, with white, pink, or red flowers, and cross-pollinating the white with the red always produces plants with pink flowers. The specific rule is that all of the first generation is pink, and when pink-flowered plans are crossed with each other, one-fourth of the offspring is white, one-fourth red, and half pink. So we can say that the inheritance mechanism works by this rule. However, no mechanism in the reproductive system keeps counting the numbers of offspring to make sure that the proportions come out right: I have made so and so many white flowers, its time to make the same number of red flowers. It is not the rule that makes the proportions come out in a certain way. The proportions are an outward reflection of the mechanism that passes traits from one generation to the next. It is significant, however, that long before chromosomes or genes, or RNA and DNA were discovered, people speculated correctly about a hereditary mechanism that would produce offspring in those proportions. Clearly, the laws provided a useful description of the behavior, and accurate description often leads to discovery and explanation.

The situation is similar with regard to language and to mental functions at large. For example, we attribute the patterns of a language to its grammar and we devise sets of rules by which the grammar works. However, it is not the grammar that generates sentences in us when we speak or write. The regularities captured in the grammar are an outward expression of our underlying mechanisms for language - the grammar is an emergent phenomenon. This distinction is easily lost when we produce language output with computers, for there we actually use the grammar to generate sentences, and we work hard to develop a comprehensive grammar for a language. And when we think of the computer as a model of the brain and use computers to model mental functions, we tacitly assume that the brain uses grammar rules to generate language. Formal logic as a model of thinking can be criticized on similar grounds. It may describe rational thought but it does not explain thinking. Our understanding of the mechanisms of mind is not yet sufficient to allow us to explain thinking and language. The best we can do is to describe them, but as our descriptions improve, our chances for discovering the mechanisms improve.

THE BRAIN AS A COMPUTER FOR MODELING THE WORLD, AND OUR MODEL OF THE BRAIN AS COMPUTING

It is useful to think of the brain as a computer if we make the analogy between the two sufficiently abstract. So what in computers should we look at? The organization of computation as a sequence of programmed instructions for manipulating pieces of data stored in memory seems like an overly restricted a model of how the brain or the mind works. A more useful analogy is made at the level of computers as state machines, the states being realized as configurations of matter, or patterns in some physical medium. Mental states and subjective experience then correspond to – or are caused by – physical states so that when a physical state repeats, the corresponding subjective experience repeats. Thus the patterns that define the states are the objective counterpart of the subjective experience. Our senses are the primary source of the patterns, and our built-in faculties for pleasure and pain give primary meaning to some of the patterns. Brains are wired for rich feedback, and when the feedback works in such a way that an experience created by the senses – i.e., a succession of states – can later be created internally, we have the basis for learning. With learning, rich networks of meaningful states can be built.

The evolutionary function of this computer is to make the world predictable: the brain models the world as the world is presented to us by our senses. It appears to compute with patterns of activity over large sets of neurons. To study such computing mathematically, we can model the patterns by large patterns of bits, emphasizing the large size of the patterns, as that gives the models their power. The key question is, how do patterns that have already been established and have become meaningful, give rise to new patterns; how do existing concepts give rise to new concepts.

I have used the binary Spatter Code (Kanerva, 1996) to model computing with large patterns. The code is related to Plates Holographic Reduced Representation (HRR; Plate, 1994) and allows simple demonstrations of it. The representation is distributed so that every item of information that is included in a composed pattern – every constituent pattern – contributes to every bit of the composed pattern: the patterns are holographic or holistic.

COMPUTING WITH LARGE PATTERNS

The following description is in traditional symbolic terms and uses a two-place relation $r(x,y)$ and a triplet $t = (x, y, z)$ as examples.

Space of Representations

All HRRs, including the Spatter Code, work with large random patterns, or high-dimensional random vectors. All things – variables, values, composed structures, mappings between structures – are elements of a common space: they are very-high-dimensional random vectors with independent, identically distributed components The dimensionality of the space, denoted by N, is usually between 1,000 and 10,000. The Spatter Code uses dense binary vectors (i.e., 0s and 1s are equally probable). The vectors are written in boldface, so that **x** stands for an N-vector representing the variable or role x, and **a** stands for an N-vector representing the value or filler a, for example.

Item Memory or Clean-up Memory

Some operations produce approximate vectors that need to be cleaned up (i.e., identified with their exact counterparts). That is done with an item memory that stores all valid vectors known to the system, and retrieves the best-matching vector when cued with a noisy vector, or retrieves nothing if the best match is no better than what results from random chance. The item memory performs a function that, at least in principle, is performed by an autoassociative neural memory.

Binding

Binding is the first level of composition in which things that are very closely associated with each other are brought together. A variable is bound to a value with a binding operator that combines the N-vectors for the variable and the value into a single N-vector for the bound pair. The Spatter Code binds with coordinatewise (bitwise) Boolean Exclusive-OR (XOR, $\otimes$), so that the variable x having the value a (i.e., $x = a$) is encoded by the N-vector $\mathbf{x}\otimes\mathbf{a}$ whose nth bit is the bitwise XOR $x_n \otimes a_n$ (x_n and a_n are the nth bits of **x** and **a**, respectively). An important property of all HRRs is that binding of two random vectors produces a random vector that resembles *neither* of the two.

Unbinding

The inverse of the binding operator breaks a bound pair into its constituents: finds the filler if the role is given, or the role if the filler is given. The XOR is its own inverse function, so that, for example, $(\mathbf{x}\otimes\mathbf{a})\otimes\mathbf{a} = \mathbf{x}$ finds the vector to which **a** is bound in $\mathbf{x}\otimes\mathbf{a}$.

Merging

Merging is the second level of composition in which identifiers and bound pairs are combined into a single item. It has also been called 'superimposing' (superposition), 'bundling', and 'chunking'. It is done by a *(normalized) mean* vector, and the merging of **G** and **H** is written as $[\mathbf{G} + \mathbf{H}]$, where $[\ldots]$ stands for normalization. The relation $r(a, b)$ can be represented by merging the representations for r, '$x = a$', and '$y = b$'. It is encoded by

$$\mathbf{R} = [\mathbf{r} + \mathbf{x}\otimes\mathbf{a} + \mathbf{y}\otimes\mathbf{b}]$$

The normalized mean of binary vectors is given by bitwise majority rule, with ties broken at random. An important property of all HRRs is that merging of two or more random vectors produces a random vector that resembles *each* of the merged vectors.

Distributivity

In all HRRs, the binding and unbinding operators distributes over the merging operator, so that, for example,

$$[\mathbf{G} + \mathbf{H} + \mathbf{I}]\otimes\mathbf{a} = [\mathbf{G}\otimes\mathbf{a} + \mathbf{H}\otimes\mathbf{a} + \mathbf{I}\otimes\mathbf{a}]$$

Distributivity is a key to analyzing HRRs.

Probing

To find out whether the vector **a** appears bound in another vector **R**, we probe **R** with **a** using the unbinding operator. For example, if **R** represents the above relation, probing it with **a** yields a vector **X** that is recognizable as **x** (**X** will retrieve **x** from the item memory). The analysis is as follows:

$$\mathbf{X} = \mathbf{R}\otimes\mathbf{a} = [\mathbf{r} + \mathbf{x}\otimes\mathbf{a} + \mathbf{y}\otimes\mathbf{b}]\otimes\mathbf{a}$$

which becomes

$$\mathbf{X} = [\mathbf{r}\otimes\mathbf{a} + (\mathbf{x}\otimes\mathbf{a})\otimes\mathbf{a} + (\mathbf{y}\otimes\mathbf{b})\otimes\mathbf{a}]$$

by distributivity and simplifies to

$$\mathbf{X} = [\mathbf{r}\otimes\mathbf{a} + \mathbf{x} + \mathbf{y}\otimes\mathbf{b}\otimes\mathbf{a}]$$

Thus **X** is similar to **x**; it is also similar to $\mathbf{r}\otimes\mathbf{a}$ and $\mathbf{y}\otimes\mathbf{b}\otimes\mathbf{a}$, but they are not stored in the item memory and thus act as random noise.

The functions described so far are sufficient for traditional symbol processing, for example, for realizing a Lisp-like list-processing system. Holistic mapping, which is discussed next, is a parallel alternative to what is traditionally accomplished with sequential search and substitution.

Holistic Mapping and Simple Analogical Retrieval

Probing is the simplest form of holistic mapping. It approximately maps a composed pattern into one of its bound constituents, as discussed above and seen in the following example. Let **F** be a holistic pattern representing France: that its capital is Paris, geographic location is Western Europe, and monetary unit is franc. Denote the patterns for capital, Paris, geographic location, Western Europe, money, and franc by **ca**, **Pa**, **ge**, **WE**, **mo**, and **fr**. France is then represented by the pattern

$$\mathbf{F} = [\mathbf{ca}\otimes\mathbf{Pa} + \mathbf{ge}\otimes\mathbf{WE} + \mathbf{mo}\otimes\mathbf{fr}]$$

Probing **F** for "the Paris of France" is done by mapping (XORing) it with **Pa** and it yields

$$\mathbf{F}\otimes\mathbf{Pa} = [\mathbf{ca} + \mathbf{ge}\otimes\mathbf{WE}\otimes\mathbf{Pa} + \mathbf{mo}\otimes\mathbf{fr}\otimes\mathbf{Pa}]$$

(see 'Probing' above) and is approximately equal to **ca**:

$$\mathbf{F}\otimes\mathbf{Pa} \approx \mathbf{ca}$$

XORing with **Pa** has mapped **F** approximately into **ca**, meaning that Paris is France's capital.

Much more than that can be done in a single mapping operation, as shown in the following two examples. Let **S** be a holistic pattern for Sweden with capital Stockholm (**St**), located in Scandinavia (**Sc**), and with monetary unit krona (**kr**). This information about Sweden is then represented by the pattern

$$\mathbf{S} = [\mathbf{ca}\otimes\mathbf{St} + \mathbf{ge}\otimes\mathbf{Sc} + \mathbf{mo}\otimes\mathbf{kr}]$$

We can now ask 'What is the Paris of Sweden?' If we take the question literally and do the mapping **S**⊗**Pa**, as above, we get nothing recognizable, so we must take Paris in a more general sense. 'Paris of France' gave us a recognizable result above (i.e., approximately **ca**), so we can use it: we can map **S** (XOR it) with **F**⊗**Pa** and we get

$$\mathbf{S}\otimes\mathbf{F}\otimes\mathbf{Pa} \approx \mathbf{St}$$

which is recognizable as the pattern for Stockholm. The derivation is based on distributivity and is similar to the one given under 'Probing'. The significant thing in **S**⊗**F**⊗**Pa** is that **S**⊗**F** can be thought of as a binding of two composed patterns of equal status, rather than a binding of a variable to a value, and also as a holistic mapping between France and Sweden, capable of answering analogy questions of the kind 'What is the Paris of Sweden?' and 'What is the krona of France?'

Holistic mapping allows *multiple substitutions* at once. What will happen to the pattern for France if we substitute Stockholm for Paris, Scandinavia for Western Europe, and krona for franc, all at once, and how is the substitution done? We create a mapping pattern as above, by binding the corresponding items to each other with XOR and by merging the results:

$$\mathbf{M} = [\mathbf{Pa}\otimes\mathbf{St} + \mathbf{WE}\otimes\mathbf{Sc} + \mathbf{fr}\otimes\mathbf{kr}]$$

Mapping the pattern for France with **M** then gives

$$\begin{aligned}
&\mathbf{F}\otimes\mathbf{M}\\
&\quad= [\mathbf{ca}\otimes\mathbf{Pa} + \mathbf{ge}\otimes\mathbf{WE} + \mathbf{mo}\otimes\mathbf{fr}]\\
&\qquad\otimes [\mathbf{Pa}\otimes\mathbf{St} + \mathbf{WE}\otimes\mathbf{Sc} + \mathbf{fr}\otimes\mathbf{kr}]\\
&\quad= [\ \mathbf{ca}\otimes\mathbf{Pa}\\
&\qquad\otimes [\mathbf{Pa}\otimes\mathbf{St} + \mathbf{WE}\otimes\mathbf{Sc} + \mathbf{fr}\otimes\mathbf{kr}]\\
&\qquad+ \mathbf{ge}\otimes\mathbf{WE}\\
&\qquad\otimes [\mathbf{Pa}\otimes\mathbf{St} + \mathbf{WE}\otimes\mathbf{Sc} + \mathbf{fr}\otimes\mathbf{kr}]\\
&\qquad+ \mathbf{mo}\otimes\mathbf{fr}\\
&\qquad\otimes [\mathbf{Pa}\otimes\mathbf{St} + \mathbf{WE}\otimes\mathbf{Sc} + \mathbf{fr}\otimes\mathbf{kr}]\ ]
\end{aligned}$$

by distributivity, which becomes

$$\begin{aligned}
&[\ [\mathbf{ca}\otimes\mathbf{Pa}\otimes\mathbf{Pa}\otimes\mathbf{St} + \mathbf{ca}\otimes\mathbf{Pa}\otimes\mathbf{WE}\otimes\mathbf{Sc}\\
&\quad + \mathbf{ca}\otimes\mathbf{Pa}\otimes\mathbf{fr}\otimes\mathbf{kr}]\\
&+ [\mathbf{ge}\otimes\mathbf{WE}\otimes\mathbf{Pa}\otimes\mathbf{St} + \mathbf{ge}\otimes\mathbf{WE}\otimes\mathbf{WE}\otimes\mathbf{Sc}\\
&\quad + \mathbf{ge}\otimes\mathbf{WE}\otimes\mathbf{fr}\otimes\mathbf{kr}]\\
&+ [\mathbf{mo}\otimes\mathbf{fr}\otimes\mathbf{Pa}\otimes\mathbf{St} + \mathbf{mo}\otimes\mathbf{fr}\otimes\mathbf{WE}\otimes\mathbf{Sc}\\
&\quad + \mathbf{mo}\otimes\mathbf{fr}\otimes\mathbf{fr}\otimes\mathbf{kr}]\ ]
\end{aligned}$$

again by distributivity. That simplifies to

$$\begin{aligned}
&[\,[\mathbf{ca}\otimes\mathbf{St} + \mathbf{ca}\otimes\mathbf{Pa}\otimes\mathbf{WE}\otimes\mathbf{Sc}\\
&\quad + \mathbf{ca}\otimes\mathbf{Pa}\otimes\mathbf{fr}\otimes\mathbf{kr}]\\
&+ [\mathbf{ge}\otimes\mathbf{WE}\otimes\mathbf{Pa}\otimes\mathbf{St} + \mathbf{ge}\otimes\mathbf{Sc}\\
&\quad + \mathbf{ge}\otimes\mathbf{WE}\otimes\mathbf{fr}\otimes\mathbf{kr}]\\
&+ [\mathbf{mo}\otimes\mathbf{fr}\otimes\mathbf{Pa}\otimes\mathbf{St} + \mathbf{mo}\otimes\mathbf{fr}\otimes\mathbf{WE}\otimes\mathbf{Sc}\\
&\quad + \mathbf{mo}\otimes\mathbf{kr}]\,]
\end{aligned}$$

and is recognizable as

$$[\mathbf{ca}\otimes\mathbf{St} + \mathbf{ge}\otimes\mathbf{Sc} + \mathbf{mo}\otimes\mathbf{kr}]$$

In other words,

$$\mathbf{F}\otimes\mathbf{M} \approx \mathbf{S}$$

so that a single mapping operation composed of multiple substitutions changes the pattern for France to an approximate pattern for Sweden, recognizable by the clean-up memory.

TOWARD A NEW MODEL OF COMPUTING

Holistic representation and holistic mapping hint at the possibility of organizing computing around analogy. However, the examples that I have shown are not very strong. This could mean that large random patterns and the suggested operations on them are not a good way to compute, but it is also possible that they are, but that we are not using them correctly. What stands out about the examples is that they are built around established notions of variable, value, property, relation, and the like. These are high-level abstractions that help us describe abstract things to each other, but they may be poor indicators of what goes on in the brain. For example, should a pattern for a variable, such as capital city in the above examples, be related to patterns that stand for individual cities, and how should those be related to the patterns for the countries they are capitals of? There are many questions to answer before we can decide about the utility or futility of computing with large patterns.

What is appealing about large random patterns is that they have rich and subtle mathematical properties, and they lend themselves to parallel computing. Furthermore, the brain's connections and patterns of activity suggest that kind of computing.

For a computer to work like the human mind, it must be extremely flexible in its use of symbols. It cannot stumble on the multiplicity of meanings that a word can have but rather it must be able to benefit from the multiplicity. The human mind conquers the unknown by making analogies to that which is known, it understands the new in terms of the old. In so doing it creates ambiguity or, rather, it creates rich networks of mental connections and becomes robust.

My hunch is that after we understand how the brain handles analogy – how it treats one thing as another – and have programmed it into computers, programming computers to handle language will be an easy task, but it will not be easy before.

REFERENCES

Kanerva, P. (1996) Binary spatter-coding of ordered *K*-tuples. In C. von der Malsburg, W. von Seelen, J.C. Vorbrüggen, and B. Sendhoff (eds.), *Artificial Neural Networks* (Proc. ICANN '96, Bochum, Germany), pp. 869 – 873. Berlin: Springer.

Plate, T.A. (1994) Distributed Representation and Nested Compositional Structure. Ph.D. Thesis. Graduate Department of Computer Science, University of Toronto.

Rumelhart, D.E, and McClelland, J.L. (1986) On learning the past tenses of English verbs. In J.L. McClelland and D.E. Rumelhart (eds.), *Parallel Distributed Processing 2: Applications*, pp. 216 – 271. Cambridge, Mass.: MIT Press.

NETAB: A NEURAL NETWORK MODEL OF ANALOGY BY DISCOVERY

J. E. McCredden

Department of Psychology
University of Queensland
jems@psy.uq.edu.au

ABSTRACT

NetAB, a two-part, three-layered feedforward neural network is used to model learning of relations and their application in discovering solutions to analogy problems. Unlike other models of analogy, NetAB allows for relations to be learned and generalised. NetAB was trained and tested against Rumelhart and Abrahamson's (1973) vector model of analogical problem solving, and against human solutions to analogy problems. Ten subjects' similarity judgements for eighteen animals were subject to multidimensional scaling, creating a conceptual space which was used both as inputs to NetAB, and for calculating solutions for the Rumelhart and Abrahamson model. The results show that while NetAB models Rumelhart et al.'s vector model favourably, neither model predicts human solutions closely. Possible reasons for the discrepancies are discussed.

Analogical reasoning is a creative thought process. Discovery by analogy occurs when a known knowledge domain (the base) is used to create concepts in a new domain (the target). The problem **a:b :: c : ?** is an example of discovery analogy in its simplest form. In this paper, we illustrate and investigate the underlying representational processes of discovery analogy, using neural networks.

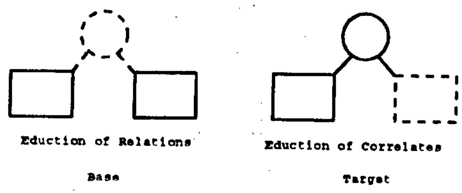


Figure 1. Eduction of Relations and Eduction of Correlates, adapted from Spearman (1923).

Spearman (1923) described analogy as comprising two components: Eduction of Relations and Eduction of Correlates (see Figure 1).

In Spearman's model, similar to the later model of Sternberg (1977), the relation between the objects in the base is not necessarily predefined, but is educed during analogical reasoning (e.g., if 'cat' and 'dog' were the two given objects in the base, then any or all of the relations 'same size', 'both friendly to man' and 'same level of domesticity' might be educed.) The educed base relation is then used actively in the target domain to educe the unknown element from the known element (e.g, if the unknown target element is 'horse', then applying any of the listed relations might educe 'cow' as the solution).

In Spearman's model the analogical process is a process of discovery, and relations are active agents in the problem solving process. We adapted Spearman's model to a connectionist framework in order to build a model of analogy based on the the three following principles:

1. Analogy by discovery rather than mapping: While much emphasis has been placed on the mapping between base and target objects and relations (Handler and Cooper, 1993; Holyoak, 1989; Forbus, Gentner, & Law, 1995; Hummel & Holyoak, in press), some models have emphasised the discovery aspects of analogy (Halford et al., 1993; Mitchell, 1990; Plate 1993). In most of these models, however, relations and arguments are pre-defined within some knowledge structure. In contrast, the current work is aimed at modelling discovery analogy where relations can be learned and generalised so that new concepts can be created via the analogy process.

2. Relations and concepts represented in similar ways: To allow structures to be represented (Gentner 1983), models must allow nesting of relations inside of higher order relations. Thus concepts and relations must be represented in similar ways so that they can be used interchangeably inside structures (see Plate, 1991).

3. Relations as active agents in processing: Classical representations for models of analogy require both a knowledge representation, and a set of processes to operate on the knowledge base. Neural networks differ from propositional representations in that representations (i) can be learned and (ii) can be active components in information processing. For example (Wiles, Stewart[1] & Bloesch, 1990) showed how every element (object or relation) input to a recurrent net is an active operator on the concept space represented by the hidden unit space (Elman, 1989). Such active representation capabilities for relations may be necessary for modelling "structural alignment" (Goldstone, 1991), which demonstrates how relations act as powerful operators in how subjects conceptualise base and target knowledge. (Gentner & Markman, 1993).

In order to model discovery analogy, relations must be modelled as active entities such that they can be (i) created in the base domain and (ii) applied in the target domain to create new concepts. Two neural net models which can be viewed as modelling these processes are HRR (Plate, 1993) and STAR (Halford et al., 1993).

Currently, there is a limited theoretical basis for describing and specifying relations both within classical and connectionist systems. In semantic nets, relations are represented as nodes, or links (Quillian, 1968), and in productions systems, they are propositions represented as role-filler structures (Anderson, 1973). In neural net models, they are sometimes treated as elements just like their object arguments (e.g., Handler and Cooper, 1993). In these models there is often a confusion between the label for a relation (e.g. 'larger-than') and the relation itself (e.g. the relation *larger-than* involving all instances of one object being larger than another). Furthermore, these models make no provision for explicit representation of the bindings of the arguments and labels into relational instances.

Representat ion of Relations

The mathematical definition of relations is used by Halford and Wilson (1982) to specify relational knowledge. For example a binary relation is specified as a subset of ordered pairs from the Cartesian product of two domains, . Given this definition, a model of relations must have the ability to represent both the overall relation and each instance . Halford et al. (1987) suggest that representations of relations must comprise a vector for each argument, a vector for the label, and a representation of the binding.

Representation of binding is central to representation of relations (see Hinton, 1986). Halford et al. (1997) provide a classification system for models of relations in terms of the type of bindings used; i.e., role-filler or argument-argument, and the type of architecture used for the binding; i.e., tensor products (Halford et al., 1993; Smolensky, 1990), convolution correlation (Plate, 1991), and synchronous oscillation (Hummel & Holyoak. in press; Shastri & Ajjanagadde, 1993).

Bindings can also be represented in the hidden layer of a neural network, (Hinton, 1986). Furthermore, Elman (1989) showed how bindings are organised into meaningful regions in recurrent networks (discovered using principle components analysis). Related work on binding and representational structure in the hidden layer of recurrent nets (Wiles, Stewart, and Bloesch, 1990) and on structure in hidden unit representations in simple feedforward nets (Wiles 1993, Wiles and Ollila, 1993) has shown that bindings can be represented in the hidden layer of a feedforward net. Furthermore, these representations provide some knowledge structure within their spatial organisation, such as hierarchies, discrete regions, and intersecting regions (Wiles, 1993b).

---

[1] The author's previous surname..

Spatial systems have been used to map conceptual similarity spaces for many domains. Multidimensional scaling techniques developed by Shepard (1962) have been used to map out spaces such as the animal knowledge domain (Henley, 1969; Rips et al., 1973). Similarity spaces provide non-propositional representations for objects, where conceptual similarity is represented by spatial distance.

Spatial systems may be extended to representations of relations as well as objects. This possibility has been exploited in the work of Rumelhart and Abrahamson (1973), who used points from Henley's (1969) animal space to represent animal arguments to the similarity relation, and the vector difference between those points to represent the similarity relation between animals. Rumelhart and Abrahamson used the relation vector in a full model of analogy, described later.

Given that (i) bindings are crucial to models of relations, (ii) that binding regions are created in the hidden unit space of feedforward nets, and that (iii) relative spatial position has been used to model relations, we suggest that perhaps a feedforward net can be used to learn bindings that represent relative spatial position of concepts from a semantic space. Furthermore, we suggest that these bindings may be utilised by a further net to solve analogy problems. We explored this possibility, using animal knowledge space and Rumelhart and Abrahamson's (1973) model of relations and analogy, to design a network architecture as follows.

The Theoretical Mechanism: Rumelhart and Abrahamson's (1973) vector model of relations and analogy was used as the theoretical basis for the network. Using Henley's three dimensional Euclidean space, Rumelhart et al. showed that analogical problem solving could be modelled using vector subtraction and addition. That is, in the problem **a : b :: c : ?** the relation between the two points in animal space (a and b) can be calculated as the vector difference between the two points. Then the vector can be applied to the known point in the target domain (c) to find the solution. Thus, Rumelhart and Abrahamson propose that the solution to an analogy (S) can be calculated as: $S = (b - a) + c$. Because the geometrical shape of this formula in Euclidean space is a parallelogram, we call it the 'parallelogram model' of analogy.

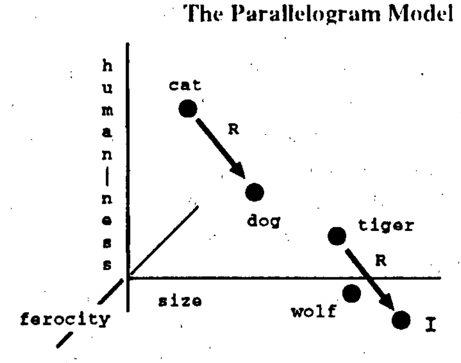


Figure 2. The parallelogram model of Rumelhart and Abrahamson as applied to the problem 'Cat is to Dog as Tiger is to What?'

Figure 2 shows the parallelogram model applied to the analogy problem 'Cat is to Dog as Tiger is to What?', where concepts are represented by coordinates in Henley's (1969) animal space. The composite relation *Dog bigger-than Cat & Dog same-ferocity as Cat & Dog less-human-than Cat* is calculated using the vector difference *Dog - Cat* and is added to the coordinates for *Tiger*, resulting in the coordinates representing the ideal solution to the analogy, *I*. Rumelhart and Abrahamson propose that the nearest animal to the ideal solution, in this case *Wolf*, is then given as the solution to the problem.

To model the parallelogram model in a network architecture, we first needed to construct animal knowledge space for a set of human subjects, to use the points in space as inputs to the net, and then obtain human solutions for analogy problems from within that space against which the net could be tested.

The Problem Domain: Human Data: Conceptual animal space was first mapped out for each subject using a similarity judgement task similar to Henley (1969). That is, for each of 18 animals chosen for the problem domain, ten subjects rank-ordered the similarity of all other animals to that animal. These judgements were then subject to multidimensional scaling, such that three dimensions emerged. Similar to Henley's dimensions, they were labelled 'size', 'ferocity', and 'domesticity'. Next, subjects were given 30 randomly created analogy problems and the solutions recorded.

A feedforward neural net, NetAB was developed to model relations over the constructed animal knowledge space, and to model analogy using the parallelogram model as the basis for the architecture. The net comprised two parts. The first part, (NetB) was designed to model the Eduction of Relations mechanism, and the second part (NetA) was designed to model Eduction of Correlates. Consequently, the experimental design for the network also had two parts. First, for each subject, conceptual space was constructed and used to train the first part of the net with relations from that subject's conceptual space. Next, the second part of the net was trained to access the relations to make identity mappings (see below). Then the Rumelhart and Abrahamson analogy solutions, the net's analogy solutions and human analogy solutions could be compared.

Using this experimental design, NetAB was designed and tested as follows:

NET*AB*: APPLICATION OF BINDING

NetAB comprised two nets: (i) NetB for representing relations, arguments, bindings of relational instances, and relation labels, and (ii) NetA for representing application of bindings, arguments, and new concepts discovered. An outline sketch of NetAB is shown in Figure 3.

In order to test the NetAB model, two criteria for correctness were used. Firstly, the parallelogram formula was used as the criterion against which to test the goodness of the NetAB model. That is, we investigated whether NetB could learn to represent relations between animal pairs (a,b) as the vector differences $(b-a)$ in hidden unit space, and whether NetA could output $S=(b-a)+c$ as the solution to the analogy. Secondly, human conceptual space was used to construct relations for training NetB, and human analogy solutions were used to test the goodness of NetAB as a model of human analogical reasoning.

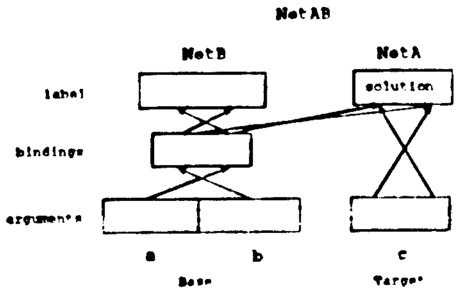


Figure 3. NetAB, comprising NetB, the binding mechanism, and NetA, the application mechanism.

Net B: The Binding Net

NetB was the Eduction of Relations net. It had six input units, six output units, and six hidden units, as follows:

Inputs: Each animal in a subject's conceptual space was represented by a vector of coordinates along each of the three dimensions. Inputs for NetB were pairs of animal vectors from a subject's conceptual space, normalised to lie in the range [+1,-1]. An example of an input set for subject1 would be 'kangaroo, koala' represented as (.71, -.43, -.71 | .51, 1.0, .28).

Outputs: Outputs were vectors representing the labels for the relations between the input pairs along each dimension (size, ferocity, and domesticity). The labels greater-than (+1, -1), equals (-1, +1), and less-than (+1, +1) were chosen such that they had no mathematical relationship to the vector difference between the two input pairs, so as to ensure symbolic, or *arbitrary* representations of labels (McCredden, 1995b). An example of an output set for subject1 corresponding to the inputs 'kangaroo, koala', would be 'greater-than (size), less-than (ferocity), less-than (domesticity)', represented as (-1 +1 | +1 +1 | +1 +1).

Hidden Units: Two dimensions (two hidden units) were required for each relational dimension (size, ferocity, domesticity), making six hidden units in total. This decision was based on previous work (McCredden, 1995a) which showed that if the hidden layer was only given one-dimension in which to represent relational bindings, the spatial location of bindings was directly related to the vector difference between the inputs, thus permitting only non-arbitrary, non-symbolic representations of relations.

Training and Testing: With 18 animals in the problem domain, and three relational dimensions for each pair of animals, there were 972 possible input-output mappings. In order to test for generalisation, the training-testing schedule chosen was to train with 70% of the input-output pairs (N=678), and to test with the other 30% of unseen relations (N=294). For each of ten subjects, NetB was run five times, using a random selection of animal pairs for training and testing. The criterion for evaluating the performance of the net was whether or not NetB's outputs were on the same side of zero as the expected output. For example, if the two inputs were 'koala, koala', the output size relation (-1, +1) i.e. 'equals' would have been expected. In this case, a NetB output of (-.02, 0.8) would have been classified as correct while an output of (.02, 0.8) would have been classified as incorrect.

Results

Table 1 shows the mean total sum of squares for NetB outputs, and the mean number of incorrect responses (for five simulations for each subject, averaged over ten subjects, rounded to integers, and converted to percentages of the total test set). The table shows that NetB learned the relations fairly well, with few errors. Further inspection of the outputs showed that most errors occured for pairs that had small differences such that NetB incorrectly classified them as equals or as a combination of either equals and less-than or of equals and greater-than. Generalisation to untrained relations was good with similar types of errors to the training set.

Test Set	TSS	s.d.	Inc.	s.d.
Trained	2	3	1	1
Untrained	16	8	5	2

Table 1. The average total sum of squares (%) and incorrect classifications (%) of relations.

NetB demonstrates how a three-layered feed-forward net is capable of learning labels for relations between pairs of animals represented by points in conceptual space. The hidden units of NetB were investigated further to see how this learning occured.

For each input pair, in order to label the relation along each dimension correctly for the outputs, the net needed only to classify the difference on the given dimension into one of three categories (greater-than, equals, or less-than). Investigations of the weights to the hidden units showed that in NetB, these categories were represented by three regions in two dimensional space. In general, each net used two hidden units to represent bindings for each of the size, ferocity, and domesticity, though there was some overlap due to correlations between relational dimensions (i.e., if animal a is larger than animal b it is often more ferocious as well.) Figure 4 depicts a case where two hidden unit dimensions coded for the size relation fairly clearly. The three binding regions created by the net for greater-than, equals, and less-than were then classified and transformed into the appropriate output labels by the hyperplanes defined by the weights and biases to the ouput units.

Figure 4 shows the greater than, less-than, and equals regions are separated and arbitrarily placed within the space. Furhter analysis of these regions for various size relations has shown that the binding regions for the small less-than and smaller greater-than relations are closer to the equals region than the binding regions for the large relations (though such a spatial layout does not always occur or is not always obvious)[2].

---

[2] The procedure for mapping out the regions in hidden unit space was repeated for several simulations until a clear hidden unit spatial representation was found. Other factors such as correlations between relational dimensions may be affecting the placement of bindings. These are currently being investigated using principle components analysis, and will be reported in future work.

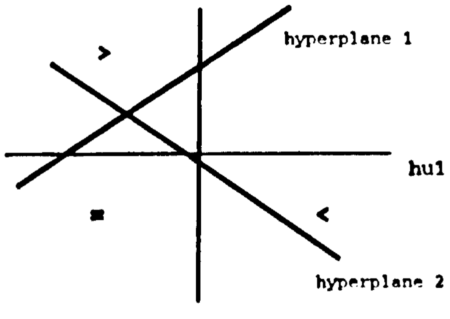


Figure 4. Binding regions for the >, =, and < relations along the size dimension for a particular run of NetB.

Further inspection of the weights and biases to the hidden units show why spatial organisations occured. In the parallelogram model, a relation between two inputs a and b is calculated as $(b-a)$. In NetB however, the bindings were calculated as some ordered measure of $(b-a)$, (denoted as $\mathrm{Ord}(b-a)$) where the absolute value of the vector difference was lost, but the relative values remained such that small vector differences gave hidden unit values closer to zero, while large vector differences gave hidden unit values closer to +/- 1. That is, the parallelogram model keeps ratio information about the relationship between two animals, while NetB keeps only ordinal information.

ANALYSIS

NetB takes perceptual-like representations as inputs and outputs symbolic labels for the relationships between inputs. NetB embodies both the animal relations as a whole, and for each instance of the relation, creates arbitrary yet information-rich bindings represented by relative position in hidden unit space. Unlike previous models of analogy the representations are learned and can generalise to unseen input pairs. Furthermore, bindings are explicitly represented during the problem, such that they can be learned and utilised by further processes. This gives NetAB an advantage over the parallelogram model where the relations are calculated and discarded.The bindings represented in NetB are used to solve analogy problems by being utilised by a further net, NetA, as described below.

NetA: The Application Net

NetA was designed to implement the second part of Spearman's model of analogy (Eduction of Correlates), and the second part of the parallelogram model, which for the analogy

a:b :: c : ? would be $S=(b-a)+c$.

Inputs: The inputs to NetA were (i) the NetB hidden unit vector representing the binding of a relation between a and b in the base, and (ii) the vector for the target animal c. An example of an input set to NetA for the 'kangaroo : koala :: zebra : ?' analogy would be the hidden unit vector for the kangaroo koala binding, (-.21, .71, -1 | 1, 1, 1), and the vector representing the target element zebra, (.89, -.57, -.62).

Outputs: The output for NetA was a three dimensional vector representing a hypothetical animal in conceptual space (Rumelhart and Abrahamson's 'ideal' solution, I.) It was assumed that some cleanup mechanism would settle on a solution which produced the animal in conceptual space closest to this point but this mechanism was not simulated. For example, in the 'kangaroo:koala :: zebra : ?' analogy, the ouput would be (-.16, .15, -.07), where the closest animal in conceptual space to this ideal solution might be 'goat'.

Training NetAB: NetB was combined with NetA to create NetAB, which was trained to map the base relation to the target domain. NetA was not trained on analogy problems, but on the simplest form of mapping; i.e., the identity relation. Thus NetAB was trained with two animal inputs to NetB, a relation label output for NetB, an animal input to NetA, and a hypothetical animal output for NetA. For example, NetAB would have been trained on (kangaroo, koala | greater-than, less-than, less-than) as inputs and outputs to NetB, and (kangaroo | koala) as input and output to NetA. NetAB was trained and tested using the same selection of training pairs used for training NetB.

Testing NetAB: After testing for the ability to map identity relations, NetAB was tested for the ability to map any relations from the base to the target. Firstly NetAB was tested with (i) the trained identity mappings (N=226) and (ii) the untrained identity mappings (N=98). Secondly, NetAB was tested for an ability to generalise the mapping task to analogy problems, so that it was tested with (iii) analogies involving relations which NetB *had* been trained with (N=30, randomly selected), (iv) analogies involving relations which NetB *had not* been trained with (N=30, randomly selected), and (v) analogies involving relations which NetB *had not* been trained with, but which humans had been given (N=30), in order to compare NetAB with human solutions.

Results

NetAB produced solutions which were points in three dimensional space, representing a hypothetical animal in a subject's conceptual space. The results for the five different tests of NetAB (where the number incorrect was the number of solutions lying more than 0.5 away from the expected solution) are shown in Table 2.

For the analogies which were presented to both humans and the net, three-way comparisons were made between NetAB solutions (AB), Rumelhart and Abrahamson's solutions (RA), and human solutions (H). The results of these comparisons are summarized in Table3 .

Using as the criterion for correctness that solutions lay within 0.5 of one another, Table 3 shows (i) how many of the solutions from each system were incorrectwhen compared with the solutions from the other systems, (ii) if thesolutions were allowed to converge to the nearest existing animals inthe space, how many were correct with respect to one another, and (iii) if the solutions were allowed to converge, how many were identical across the three systems. The results presented are the means of five simulations for each subject, averaged over ten subjects rounded to integers, and converted to percentages.

ANALYSIS

Table 3 shows that NetAB is able to utilise the bindings from NetB and apply them so as to learn the identity mapping, both for relations it has seen before and relations it has not seen before, with about a 70\% success rate. Once it has learned to map$(a \mid b)$ in the base onto$(a \mid b)$ in the target, NetAB can then do any (including analogical) mapping, and gives good results for analogies based on relations it has not been trained with, as well as on trained relations.

Test Set	av. inc.	s.d.
Trained (B) Identity	25	6
Untrained (B) Identity	34	8
Trained (B) Analogy	23	7
Untrained (B) Analogy	23	7
Human Analogy	30	10

Table 2. The average incorrect classifications (%) of identity and analogical mappings made by NetAb for relations trained and untrained by NetB.

% incorrect					
RA/AB		RA/H		AB/H	
av. 30	c.d. 10	av. 87	s.d. 7	av. 83	s.d. 7
CRA/CAB		CRA/H		CAB/H	
av. 17	s.d. 7	av . 77	s.d. 10	av. 70	s.d. 17
% identical					
av.	s.d.	av.	s.d.	av.	s.d.
13	5	4	1	3	1

Table 3. The average incorrect (%) and identical solutions (%) when the three models were compared (RA = the parallelogram model, AB = NetAB, H = Human, CRA = closest to RA solution, CAB = closest to NetAB solution).

The net's performance is good when the parallelogram solution is used as the criterion for correctness (which is not surprising since the parallelogram formula was used as the criterion for training NetB and NetAB). However, when the parallelogram model and the NetAB model are compared with human data, neither give good results. Table 3 shows that while NetAB and Rumelhart and Abrahamson solutions are similar, neither converge to solutions which are identical to human solutions very often.

Further investigation of this result was done by looking at the correlations of the base vectors (the vector difference between the target element, c and the solution for the given model) between each of the models, along each dimension (size, ferocity, and domesticity). The correlations would indicate whether the solutions for each model were alike, or very different. The correlations, averaged across all simulations, are summarized in Table 4.

The table shows that the correlations between the base vectors for the paralleogram model and for NetAB were good, middling for human data versus NetAB, and less for human data versus parallelogram solutions. In addition, the correlations were better between human data and both models for the size dimension.

The significance of the size correlations suggests a reason for the limitations of the parallelogram model (and subsequently for the NetAB model) of analogical reasoning. It could be that in human judgements, size is the most salient dimension (as illustrated by the multidimensional scaling results), and that size is often used to make judgements about the base relation in analogical reasoning regarding animals. When this occurs, all models will give similar solutions along the size dimension. However, if other dimensions, not present in the restricted three dimensional representations are used, then humans will give solutions far away from those predicted by either model. If this is the case, then both the Rumelhart and Abrahamson model and NetAB need to be adjusted so as to be able to represent relations along all conceptual dimensions and to be able to educe the salient relation from amongst all possibilities in order to apply it to the target.

RA/AB			RA/H			AB/H		
s	f	d	s	f	d	s	f	d
.9	.9	.8	.5	.3	.2	.6	.5	.4

Table 4. Correlations between the base vectors for the parallelogram model (RA), NetAB (AB), and the human data (H), averaged across all simulations for the three relational dimensions: size (s), ferocity (f), and domesticity (d).

DISCUSSION

NetAB has been used to illustrate how discovery by analogy can be viewed as comprising two component processes: Eduction of Relations and Eduction of Correlates. NetAB represents relations such that they can be learned and generalised. The model can solve analogy problems for both seen and unseen relational instances.

NetAB represents bindings in an arbitrary yet information rich concept space. Bindings are created on the run during analogy, then applied to a new domain to discover a solution to the problem. Bindings allow perceptual-like representations of pairs of animal concepts to be classified into symbolic-like categories (e.g. 'greater-than') without losing the ability to generalise to unseen instances.

While the accuracy of NetAB with respect to human solutions is limited at this stage, the processes embodied in NetAB illustrate how discovery analogy may be modelled using feed-forward nets.

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Elman, J. L. (1989). *Representation and structure in connectionist models* . Technical report 8903, Center for Research in Language, University of California, San Diego.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141-205.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Gentner, D., & Markman, A. B. (1993). Analogy: Watershed or Waterloo? Structural alignment and the development of connectionist models of analogy. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural Information systems processing systems, 5* (pp. 855-862). San Mateo, CA: Morgan Kaufmann.

Goldstone, R. L. *Similarity as structural alignment.* Research report 55, Department of Cognitive Science, Indiana University.

Halford, G. S. (1982). *The development of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Halford, G. S., Wilson, W. H., Guo, J., Gayler, R. W., Wiles, J., & Stewart, J. E. M. (1994). Connectionist implications for processing capacity limitations in analogies. In K. J. Holyoak & J. Barnden (Eds.), *Advances in connnectionist and neural computation theory, Vol. 2: Analogical connections* (pp. 363-415). Norwood, NJ: Ablex.

Handler, J. B. & Cooper, P. R. (1993). Analogical similarity: Performing structural alignment in a connectionist network. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society.* NJ: Lawrence Erlbaum Associates.

Hinton, G.E. (1986) Learning distributed representations of concepts. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society.* Amherst, MA: Lawrence Erlbaum Associates.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13(3)*, 295-355.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427-466.

McCredden, J. E. (1995) Intrinsic versus extrinsic representations of relations using neural networks. In *Proceedings of the Sixth Australian Conference on Neural Networks.* University of Sydney, Sydney, Australia.

McCredden, J. E. (1995). The components of relations illustrated in a neural network. *Paper presented at the 3rd Conference of the Australasian Cognitive Science Society*, Brisbane, Australia.

Mitchell, M. (1990) *Copycat: A computer model of high-level perception and conceptual slippage in analogy making.* PhD thesis, Computer and Communication Sciences, University of Michigan.

Plate, T. (1991, 1991 August). Holographic reduced representations: convolution algebra for compositional distributed representations. *12th International Joint Conference on Artificial Intelligence*, 30-35.

Plate, T. A. (1993). Estimating analogical similarity by vector dot-products of holographic reduced representations. In S. J. Hansen & J. D. Cowan (Eds.), *Advances in Neural Information Processing Systems 6.* San Maeto, CA: Morgan Kaufman.

Quillian, M. (1968) Semantic Memory. In M. Minsky (Ed.) *Semantic Information Processing*, Cambridge, MA: MIT Press.

Medin, D. L. & Gentner, D. (1991). Relational similarity and the non-independence of features in similarity judgements. *Cognitive Psychology*, 222-264

Rips, L. J. & Schoben, E. J. (1973). Semantic Distance and the verification of semantic relations *Journal of Verbal Learning and Verbal Behaviour 12*, 1-20.

Rumelhart, D. E. & Abrahamson, A. A. (1973). A model for analogical reasoning. *Cognitive Psychology 5*, 1-28.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences, 16(3)*, 417-494.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence, 46(1-2)*, 159-216.

Spearman, C. E. (1923). *The nature of intelligence and the principles of cognition.* London: MacMillan.

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review, 31*, 356-378.

Wiles, J. (1993). Represenation of variables and their values in neural networks. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society.* NJ: Lawrence Erlbaum Associates.

Wiles, J. (1993). Intersecting regions: The key to combinatorial structure in hidden unit space. In S. J. Hansen & J. D. Cowan (Eds.), *Advances in Neural Information Processing Systems 6*. San Maeto, CA: Morgan Kaufman.

Wiles, J., Stewart, J.E.M., & Bloesch, A. (1990). *Patterns of activation are operators in recurrent networks.* Technical report 189, Department of Computer Science, University of Queensland, Australia.

CONNECTIONS, BINDING, UNIFICATION AND ANALOGICAL PROMISCUITY

Ross W. Gayler, Roger Wales
Department of Psychology, The University of Melbourne
Parkville VIC 3052, AUSTRALIA
r.gayler@psych.unimelb.edu.au; r.wales@psych.unimelb.edu.au

ABSTRACT

This paper claims that higher cognition implemented by a connectionist system will be essentially analogical, with analogical mapping by continuous systematic substitution as the core cognitive process. The centrality of analogy is argued to be necessary in order for a connectionist system to use representations that are effectively symbolic. In turn, these representations are argued to be a necessary consequence of a sequence of broad design decisions needed to address technical problems in adapting a connectionist system for higher cognition. The design decisions are driven by the demands of a paradigmatic cognitive task and the desire to remain faithful to the constraints of connectionist components. Thus, the argument explains the origin of symbolic representations and analogy as necessary consequences of task demands and connectionist processing capabilities.

INTRODUCTION

One of the more persistent problems in cognitive science is the reconciliation of the emergent functional properties of human cognition with the apparently much more limited functional capabilities of the neural systems that implement them. Higher cognition has been most successfully modelled in terms of symbolic computations that appear implausibly difficult to implement neurally. On the other hand, connectionist systems (the currently favoured paradigm for modelling the presumed computational processes of neural systems) appear to be neurally implementable but far less capable than symbolic systems of implementing the desired cognitive functions.

Despite the relative success of symbolic computation the shortcomings of classical Artificial Intelligence (based on symbolic computation) suggest that simply scaling up the size and speed of current symbolic systems will not yield the desired cognitive functions. One response to this situation is to focus on building connectionist systems. This action is based on taking human cognition as an existence proof for the possibility of implementing higher cognition with a connectionist system.

Some researchers have implemented classic symbolic architectures in connectionist systems. For example, Touretzky and Hinton (1988) built a Distributed Connectionist Production System. We have chosen not to follow this approach of implementing known symbolic processes because we believe it will be bound by the limitations of current symbolic models.

The problem of attempting to find a connectionist architecture with the desired cognitive properties can be cast as one of efficiently searching design space. Given the vast number of potential connectionist systems we need a strategy to guide our exploration of designs. We have chosen to be guided by the constraints imposed by connectionist computational elements and the problems to be solved by higher cognition. By remaining true to the connectionist raw material we hope to allow solutions that are obscured by taking symbolic operations as the primitive functions of processing. If it turns out that the emergent properties of such a connectionist system may be characterised as symbolic, then that is further evidence for the plausibility of the system (given the success of symbolic models - but not symbolic implementations).

In this paper our strategy for exploration of design space may be summarised by the question: Starting with a simple connectionist system; what minimal design decisions might lead to a capability for higher cognition?

This question arises from an evolutionary stance. It is taken as given that higher cognition is the function of an artefact built from neural components and designed through evolution to solve certain survival problems. Given the conservative nature of evolutionary design and that connectionism is an appropriate model of neural computation we believe that a sequence of minimal modifications starting from the simplest connectionist system will be sufficient to yield a system with the desired cognitive capabilities.

This paper suggests a series of design problems and broad design approaches to their solution. (We aspire to precise, implementable design choices, but that is work in progress.) Since the modifications to the connectionist architecture are constrained to be minimal, the hope is that the resultant architecture will be practically implementable. Furthermore, we argue that the symbolic properties of higher cognition and the centrality of analogy to cognition arise as necessary consequences of the design decisions motivated by connectionist problems. Thus, the argument (to the extent it is successful) explains the emergence of symbolic properties and the centrality of analogy.

THE DESIGN PROBLEM

In order to specify the design problem that is the basis of this argument it is necessary to state the functional capabilities that are required of the final system and the design of the initial connectionist system.

REQUIRED FUNCTIONAL CAPABILITIES

Specifying higher cognition in entirety is obviously too ambitious a sub-goal for this paper. The critical cognitive characteristics sought are embodied in the ability to follow novel instructions. To make this concrete we will settle on an arbitrary but paradigmatic task[1] to be carried out by the cognitive system. The task is to be able to follow novel instructions, such as:

I will show you an artificially coloured picture of an animal and play the sound of an animal. If the sound belongs to the pictured animal you must name the colour of the pictured animal, otherwise name the animal that made the sound.

For the purposes of this paper the issues of language understanding required for comprehension of the instruction are ignored and we focus on complying with the instruction once it has been comprehended.

INITIAL CONNECTIONIST SYSTEM

The system is to be implemented with typical connectionist units. That is, each unit may have multiple inputs and a single output. All inputs and outputs are to be graded, scalar quantities. The output is a nonlinear monotonic function of the weighted sum of the inputs or products of groups of the inputs.

The system is to have a fixed architecture. That is, the pattern of interconnection of units is not to vary or be constructed as a function of the current task. For example, this rules out the ACME model of analogical mapping (Holyoak & Thagard, 1989) as a permissible architecture because the neural net is constructed specifically for each problem.

The final constraint on the connectionist architecture arises from the nature of the task. The system is to implement the "top" level of cognition. Therefore, it must be capable of integrating multiple sensory modalities. We assume that the full system will have other levels where the sensory modalities are processed separately. The boundary of the system of interest is to be expanded until it reaches the point at which the modalities are separately represented.

---

[1] This paper was inspired by Hadley (1998). He used an example of following novel instructions to motivate his argument that most human mental skills must reside in separate connectionist modules and thereby instantiate a "classical architecture".

REPRESENTATIONAL DESIGN DECISIONS

REPRESENTING NOVEL CONCEPTS

The first design problem to address is representational flexibility. The paradigmatic task requires the creation of new concepts. At the very least it requires a concept of the instruction to be followed. Therefore, the system must be capable of representing arbitrary new concepts[2].

This constraint rules out local representations where each unit has a fixed meaning. If all the units are pre-allocated to concepts how can new concepts be represented? It also seems implausibly wasteful to have unallocated units waiting to be allocated to what may be ephemeral concepts.

This constraint can be avoided by using distributed representations (that is, the representation consists of the pattern of activities across the units). However, not all distributed representations avoid the problem. The same argument would apply to distributed representations where the individual units have fixed meanings as features. Any fixed allocation of meanings to units will limit the representational possibilities.

A related argument comes from the requirement that the system should integrate information from multiple modalities. Below the point of integration the information from separate modalities travels on separate pathways. Above the point of integration it would be possible to have disjoint segments of the representation devoted to different modalities, but this would be wasteful of representational resources (units).

It is possible to have information from separate modalities represented over the same units at different times provided that context information is available as part of the representation. This will allow the representation to be interpreted differently depending on the source.

This style of representation results in the units having context dependent meanings. The activities of individual units become meaningless unless they are able to be interpreted in the context of the activities in the other units in the representation[3].

Therefore, the first design decision is to use a distributed representation where the individual units do not have fixed meanings[4]. Kanerva (1995) has also argued that fixed feature representations are impractical for open-ended domains.

IMMEDIATE LEARNING

The next design problem arises from the need for immediate learning. The system must be able to learn[5] the novel concepts immediately from a single exposure. This rules out iterative weight adjustment techniques such as backpropagation because they are too slow, typically requiring thousands of exposures[6].

What we want is that some specific output pattern (representing the novel concept) should be produced in response to a specific combination of input patterns. This is equivalent to saying that we want to associate the input and output patterns and to retrieve the output pattern given the inputs as a cue.

---

[2] Any use of "concept" and "representation" begs many philosophical questions. For current purposes read "concept" as "a mental state" and "representation" as "a physical state standing for a mental state".

[3] Context dependency does not have to be all or none. At one end of the scale we can put representations where the unit activities can be interpreted in isolation. At the other end of the scale we have representations in which only the entire pattern of activations has significance. Between these extremes are representations where subsets of the activation pattern may be assigned meanings.

[4] The degree of context dependency involves tradeoffs. At the context-independent end we restrict the representational capacity and flexibility. At the total pattern end any corruption of the pattern would completely change the meaning. We suspect that a good trade-off might exist not too far from the context-independent end of the scale where each unit is interpretable in the context of a small number of other units (relative to the total number of units) and participates in multiple, overlapping, meaningful subpatterns.

This can be achieved with binding operators for associating patterns and unbinding operators for retrieving components from bound patterns. Generically, if b=bind(x,y) then unbind(b,x)=y where b, x, and y are patterns. A variety of binding operators have been developed (Gayler, 1998; Kanerva, 1996; Plate, 1994; Smolensky, 1990)[7]. All the binding and unbinding operators are able to be implemented as connectionist primitives able to operate in a single time step.

In the example above the patterns x and y may be taken as input and output patterns respectively. The pattern b (which is able to be created in a single time step) represents the association of x and y. If b is present[8] in an environment where unbinding occurs automatically, the presentation of the input pattern x will result in the creation of the output pattern y.

The corresponding design decision is that the connectionist system should implement immediate learning as pattern association via bind and unbind operators.

COMPATIBILITY OF REPRESENTATIONS

The next problem arises because the paradigmatic task requires close interaction of short term and long term knowledge. The task calls on pre-existing skills such as animal identification and requires their integration with the short term concepts of the task and the work in progress. Therefore, there is a requirement that the system is able to integrate short term and long term knowledge.

In a traditional connectionist system short term knowledge is usually implemented as activations of units and long term knowledge as connection weights. This implementation captures the relative persistence of the two types of knowledge. However, as abstract representations, the activation vector and weight matrix are incommensurable and can only indirectly influence each other via the processing function of the system. Our intuition is that the use of such different representations for short and long term knowledge will make integration difficult.

Therefore, the next design decision is to require short and long term knowledge to be represented (though not necessarily implemented) in the same way. That is, bindings in short and long term memory must have identical representations and have identical effects on operations.

These properties automatically come from binding methods that are based on element-wise multiplication operations of terms. In connectionist systems activations and weights interact multiplicatively. Therefore, in a binding operation short term knowledge (activations) and long term knowledge (weights) may be used interchangeably provided that they are all represented as vectors of the same dimension.

A newly created binding may be kept as an activation pattern or added into a weight vector with equivalent effect. Thus short and long term knowledge are identical in terms of their ability to be interrogated by current processing. The only asymmetry between the two storage forms is that whereas activation patterns may interact with other activation patterns and weight patterns, weight patterns may

---

[5] By learning we mean changing the state of the system so that future occurrences of the novel concept are recognised. This changed state must be able to persist longer than the immediate span of attention.

[6] This is not to say that iterative weight adjustment procedures have no place in connectionist systems, only that they are inadequate for short term cognitive learning.

[7] Terminology differs between systems and there is some scope for confusion as each system has at least two distinct operators that might be called binding. We use the term "binding" for what might best be called structural binding. In this case each of the components is structurally required (also called role/filler binding or attribute/value binding). We use the term "bundling" for what might be called decorative binding. In this case each of the components is optional (for example, the slots of a frame). This is called superposition or chunking. (The latter term is used differently by different authors.)

[8] Binding addresses the issue of creating the association. Other connectionist mechanisms, able to operate in a single time step, exist to make the representation of the association persistent.

not interact with other weight patterns except via the mediation of an interaction with an activation pattern.

DISCUSSION OF REPRESENTATIONS

BINDINGS AS VIRTUAL NEURONS

Historically, single cell recording allowed neurophysiologists to identify stimuli that caused a neuron to fire. Over time researchers discovered cells responsive to progressively more complex stimuli. This was paralleled in the psychophysical literature by hypotheses of progressively more complex feature detectors. This led to the infamous "grandmother cell" as a *reductio ad absurdum* argument against complex feature detectors.

It seems implausible and wasteful that the brain might come prestocked with feature detectors for every concept that a person might possibly encounter (let alone what our descendants might encounter). This problem might be avoided if feature detectors could be created instantly on demand. This would be very difficult to achieve with real neurons but would be feasible if the feature detectors were virtual neurons implemented on a fixed neural base.

The design decisions so far are: that concepts will be represented by distributed patterns with individual units having no specific meaning; that association of concepts will be carried out by binding operators capable of immediate learning; and that short and long term representations of bindings will be equivalent, differing only in their form of storage and ability to interact with other bindings. These bindings may be thought of as virtual neurons. They are like neurons because for each binding an output pattern[9] may be retrieved by presentation of the associated input pattern. They are virtual because any number of neurons (bindings) may be implemented in a fixed number of real neurons or connectionist units (subject to soft capacity constraints).

The advantages of virtual neurons implemented as bindings are: that these neurons can be created on demand to represent novel concepts; that they can be created in a single exposure; that they may be ephemeral or permanent depending on whether they exist only as activation vectors or are stored in long term memory as weight changes.

BINDINGS AND SYSTEMATICITY

Fodor and Pylyshyn (1988) argued that one of the hallmarks of cognition that is explicable by a symbolic approach but not by connectionist models is systematicity. This is the property that having the capability to represent some concepts necessarily entails the capability to represent other related concepts. For example, they argued that a cognitive system able to think **Mary loves John** must necessarily be able to think **John loves Mary**.

Bindings automatically provide systematicity. If two patterns are bound together either may be used as a cue for the retrieval of the other. The notions of input and output, which are meaningful for real neurons, are not relevant to bindings as virtual neurons. In effect, whenever a binding implements a virtual neuron mapping x to y it also necessarily implements a virtual neuron mapping y to x.

It could be objected that producing all mappings between the bound patterns is not necessarily desirable. We agree, but discount this objection for two reasons. Firstly, we are concerned with higher cognitive functions. At this level we expect the major demand to be maximal exploitation of the available knowledge. Systematicity is a mechanism for generating hypotheses from prior knowledge to as yet unencountered situations. Halford (1996) has also argued for this capability which he labels "omni-directional access".

The second reason is that we have assumed that where it is important to limit the hypotheses generated by systematicity this will be achieved by the details of the representational scheme. For example,

---

[9] Note that now the output is a pattern rather than a scalar value. This is a consequence of distributed representation rather than virtualisation. In a distributed representation the vector of scalar outputs of a group of units is the more convenient level of analysis of output.

Smolensky (1990) proposed a representational scheme in which fillers were bound to roles rather than directly to each other. For example, **John loves Mary** might be represented as **bundle(bind(lover,John), bind(loved,Mary))** where **lover** and **loved** are roles and **John** and **Mary** are fillers, rather than **bind(loves,John,Mary)** where **loves**, **John** and **Mary** are all fillers. In Smolenskys representation the systematicity would be expressed with respect to role/filler pairs rather than directly between fillers.

BINDINGS AS RULES ON CONSTANTS

We have argued that bindings may be viewed as implementing virtual neurons. They may also be viewed as implementing rules or productions restricted to literal constants. A binding of x with y may be construed as implementing the rules IF x THEN y and IF y THEN x. Similarly for higher order bindings of x, y, and z: IF bind(x,y) THEN z IF bind(x,z) THEN y, IF x THEN bind(y,z), and so on. The restriction on the rules is that the antecedent and consequent consist only of literal constants because the bindings are between constant pattern vectors.

However, bindings do implement an extension relative to traditional symbolic rules. Because bindings are represented as pattern vectors they acquire some properties from vector arithmetic. The pattern vectors may be multiplicatively scaled and added. Thus it makes sense to talk about a binding operating on a mixture of patterns. In most binding systems bind(x,y+z) = bind(x,y) + bind(x,z). It is also possible to have graded similarity between vectors. This allows rules implemented as bindings to operate in a graded fashion.

PROCESSING DESIGN DECISIONS

GENERATE DISSIMILAR VECTORS

The next design decision is required as a consequence of an earlier decision. Recall that we decided to use distributed pattern vectors to represent new concepts. Every time a new concept is created a new pattern will be required to represent it. What constraints might exist on the patterns that can be used for new concepts?

Thinking of the binding as a virtual neuron there will be one or more input patterns to be associated with the output pattern representing the novel concept. The binding process imposes no constraint on the choice of output vector because any vectors may be bound[10]. The major constraint is that novel concepts require novel concept vectors. They should not be identical to any pre-existing concept vector otherwise the combination of inputs will be bound to a pre-existing concept.

In standard connectionist models much use is made of the fact that there is a graded similarity relation between vectors. To the extent that one vector is similar to another it is able to stand in for the other vector in further processing. If a new concept is truly novel the pattern representing it must not be similar to any pre-existing vectors in order to avoid having effects similar to some pre-existing concept.

Given immediate learning, pattern vectors for new concepts will be created at a point when the system is in a state of ignorance about the potential relationships between the new concept and any pre-existing concepts. Thus, even if the new concept vector should ideally be similar to some pre-existing concept vector it must necessarily be created dissimilar to all pre-existing concept vectors.

The corresponding design decision is that vectors representing new concepts should be generated to have zero or minimal similarity to all pre-existing concept vectors. It will simply be assumed that such a generation mechanism is feasible.

Vector Generation Mechanisms

We discuss some possible generation mechanisms with no particular commitment to any of them. The least interesting possibility for creating dissimilar vectors is to rely on random noise in the system. Random high dimensional vectors have close to zero similarity on average. An extra mechanism might be required for those rare occasions when the new vector happened to be similar to an old vector.

---

[10] There may be some exceptions. In multiplicative binding (Gayler, 1998) a pattern may not be bound to itself. This is not a problem in the current case because it implies that the output vector is identical to one of the input vectors (in which case the output vector has already been assigned to another concept).

Another mechanism (possibly an implementation of the previous one) is to modify a system with recurrent dynamics, similar to the Brain-State-in-a-Box (Anderson, Silverstein, Ritz, & Jones, 1977), such that pre-existing vectors become repellors in the state space. When presented with a novel stimulus the system would settle into a state different from any previously encountered state. The randomness in the choice of the new pattern might come from chaotic dynamics or the amplification of innate noise.

The most interesting possibility is that the novel concept vector might be created as a side effect of binding. For example, the pattern representing the binding (or some deterministic function of the binding pattern) of the inputs could be used as the output pattern. Multiplicative binding (Gayler, 1998) is essentially a randomising operation. The representation of bind(x,y) is not similar to either x or y when x and y are dissimilar (as we argue they should be if they represent concepts of higher cognition)[11]. Any novel combination of representations to be bound will necessarily generate a binding that is novel. This avoids the system having to decide when to create a new representation. If the combination is novel the binding and the output pattern will also be novel.

Vectors and Classical Symbols

The decision to represent novel concepts with vectors dissimilar to all pre-existing vectors leads to discrete representations. In general, two vector representations will either be identical or dissimilar[12].

The ability to bind arbitrary input and output representations means that the output representations can have arbitrary referents. The actual arbitrariness of the linkages will be guaranteed to the extent that representations for new concepts are generated at random.

Classical symbols are discrete and arbitrary, in that two symbols are either the same or different and that the form of a symbol has no necessary dependence on the referent of the symbol. Thus, the design decisions so far have led to connectionist representations that behave like classical symbols.

SYSTEMATIC VECTOR SUBSTITUTION

The design decisions taken so far have generated a new problem. If the representations of concepts are primarily dissimilar and arbitrary how can they be used? Traditional connectionist processing relies on the similarity of vector representations, which has been removed by the design decisions.

If the individual, isolated, pattern vectors do not carry information, what does? We believe that the information must be carried in the structural interrelationships of the bindings. During any episode the system will be creating many bindings which will be interrelated by the individual pattern vectors they have in common[13]. On a subsequent occasion the new episode will be recognised as equivalent to the previous episode if the structural interrelations of the bindings are the same (even though the pattern vectors composing the bindings may differ). This structural equivalence is proved if a systematic substitution of pattern vectors in the current episode yields the previous trace[14].

---

[11] In all the binding systems mentioned earlier (Gayler, 1998; Kanerva, 1996; Plate, 1994; Smolensky, 1990) the bind() operator reduces the similarity of compounds compared to their components. Considering the limiting case, the similarity of bind(x,y) to x and y is zero when the similarity of x and y is zero. Thus the overall effect of binding is to make representations less similar and the introduction of a single dissimilar pattern will render dissimilar every binding in which it participates. We are not denying the importance of those occasions when patterns and their bindings are similar, rather we are focusing on the occasions of dissimilarity because we believe they are more relevant to higher cognition and have been ignored by connectionists.

[12] It is possible for two representations to be similar and for processing to depend on that similarity. However, in a high dimensional vector space the proportion of vectors similar to any given vector becomes very small. Most pairs of vectors chosen at random will be dissimilar.

The corresponding design decision is that the basic mode of processing of the connectionist system should be continuous systematic substitution of representations for the elaboration of the currently active representations from previous representations.

Systematic substitution is relatively simple with the binding systems mentioned earlier. For example, in multiplicative binding, binding any structure to bind(a,b) will result in replacing all occurrences of a in the structure with b (and b with a)[15].

Rules with Variables

An interesting consequence of this decision is that it turns every constant pattern vector into a variable because every constant may be systematically substituted. Earlier it was noted that bindings could be thought of as rules restricted to constant terms. Without changing the representation systematic substitution allows every binding to function as a rule with variables. Thus, any encoding from any episode (even if encountered only once) becomes available as a generalised rule to the extent that it can be unified by systematic substitution with other structures.

It might be the case that in most circumstances systematic substitution is not needed because there is literal similarity between the representations. Processing based on literal similarity is equivalent to systematically substituting each pattern vector for itself. Thus, it could be the case that systematic substitution is continually occurring in the connectionist system but we can only detect it when we are operating in domains where literal similarity is not available.

Unification and Analogical Mapping

This design decision asserts that systematic substitution is a necessary component of the cognitive process because of the consequences of the earlier design decisions. Systematic substitution is at the heart of analogical mapping. Therefore, we are asserting that analogical mapping is a necessary component of the cognitive process.

It is worth expanding on the possibility that analogical mapping may be at the heart of cognition. We referred earlier to the problem of the system knowing when to create novel representations and suggested that one possibility is that the representations are functionally dependent on the inputs combined. That is, novel combinations of inputs would result in novel representations.

Given that very few situations are identical (especially if you consider the goals of the cognitive system as a representable component of the situation) this has the potential to make every representation a novel representation. The mechanism proposed here to overcome this is to use a continuous process of systematic substitution to unify[16] the current representation with all previously encountered representations.

We also referred earlier to the multiplexing of information from multiple modalities over the same representational resources. This imposed a requirement for context to be represented as a component of the content and made the interpretation of representations context dependent. As the number of contextual states increases this also has the effect of turning each representation into a novel representation (even if the entity being represented remains constant).

---

[13] A representation **bundle(prop(loves,Chris,Pat), prop(loves,Pat,Robin), prop(loves,Robin,Chris))** would be structurally equivalent to any representation of the form **bundle(prop(l,c,p), prop(l,p,r), prop(l,r,c))**.

[14] Given the episode **bundle(prop(loves,Alex,Jerry), prop(loves,Jerry,Lee), prop(loves,Lee,Alex))** the systematic substitutions {Alex⇨Chris, Jerry⇨Pat, Lee⇨Robin} would transform it into the previous episode. Equivalently, the previous episode could be transformed to the current episode by systematic substitution. For current purposes, we ignore the representational details governing whether substitution for the relation (**loves**) is allowed and whether roles and fillers might be interchangeable.

[15] For other substitution mechanisms based on binding see Halford, Wilson, Guo, Gayler, Wiles & Stewart (1994), Kanerva (1997), and Plate (1997).

[16] Unification is a proof technique used in logic programming which uses substitution of variables to make terms equivalent It is computationally expensive when implemented by a symbolic algorithm. Weber (1992) developed a connectionist system (not based on the binding techniques discussed here) that implements unification in constant time

Therefore, it is possible that the cognitive system is thoroughly promiscuous in the generation of new representations and that these representations will appear to be random when viewed in isolation. How are these novel representations to be interpreted and acted upon if they appear random?

A similar problem arises in the technicalities of vector systems. Any vector can be decomposed into contributions from a set of basis vectors. By requiring new concept representations to be arbitrary and dissimilar, we have removed from the underlying hardware the possibility of having a distinguished basis set (fixed features) for decomposing composite structures (because the basis vectors are now the concept vectors which are not known until they are created).

When faced with a pattern vector, how does the system know whether it represents a new concept or some composite structure that may be decomposed into other representations? This is important because a composite structure needs to be exploited by integrating it with related knowledge, whereas a novel concept should not be spuriously integrated with prior knowledge.

The solution proposed is to decompose it with respect to the other structures that already exist in short and long term memory. Our intuition is that this might be carried out as a continuous process of activation spreading from the active representations in short term memory, through the inactive representations in long term memory, creating further activations in short term memory.

If the process of propagating activation simultaneously pursues many systematic substitutions a shower of new activations will be created. Those activations that are identical with or consistently extend the pre-existing activations will reinforce those patterns and themselves, while inconsistent mappings die out as noise or remain as suppressed alternative decompositions to pursue if the current one becomes inconsistent.

This process can be viewed as a competition between potential decompositions. The first and most consistent decomposition would be more successful than competing decompositions at creating the feedback activations to reinforce itself. Thus the decomposition that occurs is the (or a) correct interpretation of the structure (by virtue of its success). Other potential decompositions are incorrect interpretations of the structure because they were less successful at integrating the active representations and long term memory. This automatically achieves the desired result that a concept vector should be decomposed and integrated where ever possible.

CONCLUSION

We have suggested a series of connectionist design decisions that seem to follow naturally from the nature of the cognitive task by respecting the essence of connectionist computation. These decisions[17] are:

- use distributed representations where the individual units do not have fixed meanings;
- implement immediate learning as pattern association via bind and unbind operators;
- bindings in short and long term memory must have identical representations;
- vectors representing new concepts should be dissimilar to all pre-existing concept vectors;
- the basic mode of processing should be continuous systematic substitution of pattern vectors.

As necessary consequences of these decisions the connectionist system will demonstrate behaviour typical of classical symbolic systems and place analogy as the primary cognitive process. Interpretation of these decisions suggests that cognition is promiscuously analogical and that the basic mechanism of cognition consists of a continuous process of unification through systematic substitution of currently active representations with each other, all representations in long term memory, and new representations being created. In effect, this is a continuous data mining operation on a massive scale.

---

[17] Connectionist systems exist demonstrating all but the last decision (which we believe is plausible within the current state of the art of systematic substitution by binding and unbinding). Even with an implementation of continuous systematic substitution all the design decisions will need to be integrated and there will be auxiliary problems to be solved before a connectionist model of higher cognition can be demonstrated.

REFERENCES

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413-451

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3-71.

Gayler, R. W. (1998, July). *Multiplicative binding, representation operators, and analogy.* Poster session presented at Analogy'98, Sofia, Bulgaria.

Hadley, R. F. (1998). *Connectionism and novel combinations of skills: Implications for cognitive architecture* (Technical Report SFU CMPT TR 1998-01). Burnaby, Canada: Simon Fraser University, School of Computing Science.

Halford, G. S. (1996) Relational knowledge in higher cognitive processes. In *Relational knowledge in higher cognitive processes.* Symposium conducted at the XIVth Biennial Meeting of the International Society for the Study of Behavioral Development, Quebec City.

Halford, G. S., Wilson, W. H., Guo, J., Gayler, R. W., Wiles, J., & Stewart, J. E. M. (1994). Connectionist Implications for Processing Capacity Limitations in Analogies. In K. J. Holyoak & J. A. Barnden (Eds.), *Advances in connectionist and neural computation theory, Vol. 2: Analogical connections.* Norwood, NJ: Ablex.

Holyoak, K. J. & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*, 295-355.

Kanerva, P. (1995). A family of binary spatter codes. In F. Fogelman-Soulié & P. Gallineri (Eds.), *ICANN '95, Proceedings International Conference on Artificial Neural Networks, Vol. 1* (pp. 517-522). Paris, France: EC2 & Cie.

Kanerva, P. (1996). Binary spatter-coding of ordered *K*-tuples. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, & B. Sendhoff (Eds.), *Artificial Neural Networks - ICANN 96* Proceedings 1996 International Conference, Bochum, Germany; *Lecture Notes in Computer Science* 1112), 869-873. Berlin: Springer.

Kanerva, P. (1997). Fully distributed representation. In *Proceedings RWC Symposium 1997 (Tokyo, Japan)*, 358-365.

Plate, T. A. (1994). *Distributed representations and nested compositional structure.* Ph.D. thesis, Department of Computer Science, University of Toronto.

Plate, T. A. (1997). Structure matching and transformation with distributed representations. In R. Sun & F. Alexandre (Eds.), *Connectionist-Symbolic Integration.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*, 159-216.

Touretzky, D. S., & Hinton, G. E. (1988). A distributed connectionist production system. *Cognitive Science*, *12*, 423-466.

Weber, V. (1992). Connectionist unification with a distributed representation. In *International Joint Conference on Neural Networks IJCNN' 92, Beijing, China*, 555-560.

STRUCTURE AND PRAGMATICS IN ANALOGICAL INFERENCE

Arthur B. Markman
Department of Psychology, University of Texas
Mezes Hall 330, Austin, TX 78712, markman@psy.utexas.edu

Adalis Sanchez
Department of Psychology, Columbia University
406 Schermerhorn Hall, New York, NY 10027

ABSTRACT

An important aspect of the process of forming analogies is the ability to extend knowledge of a target domain by virtue of its similarity to a base domain. Extant theories of analogy suggest that information is carried from base to target when it is connected to a correspondence between the domains and is structurally consistent with the current match. Some theories further suggest that information is likely to be carried from base to target when it is pragmatically relevant to the current situation. I present studies that examine the contributions of structure and pragmatic relevance on analogical inference using a technique in which people play the role of a student or a financial officer transferring from one college to another. The results indicate that systematicity and pragmatic relevance play distinct roles in analogical inference.

INTRODUCTION

There is general agreement among researchers that analogy involves sub-processes including representing the domains, finding a mapping between them, verifying the goodness of the mapping, and carrying inferences from one domain (called the *base*) to a second domain (called the *target*). The process of analogical inference has been the object of study for two reasons. First, the ability to create analogical inferences is an important avenue of knowledge change, because it allows one domain to be extended by virtue of its similarity to another. Second, existing computational models of analogy disagree on the mechanisms by which candidate inferences are generated, and so data that bear on this issue will help constrain these computational models.

Much of the work related to analogical inference has been done in the context of transfer in problem solving (e.g., Gick & Holyoak, 1980; Ross, 1989). This work has focused primarily on how whole solutions to old problems can be carried over to new ones. More recently, work has focused on factors that determine which pieces of information about a base domain are likely to be inferred of a target. Two central constraints on inference that have been studied are systematicity (Clement & Gentner, 1991; Markman, 1997). and pragmatics (Spellman & Holyoak, 1996). In this paper, I first briefly review the work on systematicity and pragmatics. Then, I present three studies that examine both pragmatics and systematicity in order to examine their relative importance as constraints on inference. I conclude with a discussion of the implications of this work for existing models of analogy.

SYSTEMATICITY AND PRAGMATICS

Analogical inference must be constrained, because not every fact true of a base domain will also be true of the target. Indeed, for distant analogues, most of the facts about the base domain will not be true of the target. If every fact about the base were carried to the target, then most of the information inferred would be false, and the inference process would not be useful (because the reasoner would waste considerable time rejecting false inferences).

The first constraint—systematicity—is the notion that connected relational systems are preferred to collections of individual relations (Gentner, 1983; Gentner, 1989). Systematicity constrains inference by requiring that the facts carried over from the base be connected to matching information in the target. That is, the inferences must involve *shared system* facts. The assumption is that a fact connected to a match between base and target is more likely to be relevant than is a fact not connected to the match.

For example, imagine that you know two facts about a friend:

(1) John likes to eat ice cream *causing* him to be slightly overweight.

(2) John likes old movies *causing* him to stay up late watching TV.

Suppose that you then strike up an email correspondence with a new person, Mary, who likes old movies. Systematicity suggests that you should infer that Mary stays up late watching TV (a shared system fact) rather than that Mary is slightly overweight (a *nonshared system fact*). Previous studies of analogical inference have shown that people making analogical inferences are much more likely to infer shared system facts than nonshared system facts (Clement & Gentner, 1991; Markman, 1997).

In addition to systematicity, pragmatic information also seems useful for constraining analogical inference. If you know in advance that a particular piece of information is of interest, then you should be more likely to infer that information. For example, if you strike up an email correspondence with Mary, and realize that she generally reminds you of your friend John, then you might want to make inferences about Mary based on what you know about John (Andersen & Cole, 1990). If you are particularly interested in whether Mary watches TV, you might focus selectively on the inference that she stays up late watching TV, because it is relevant to your goals.

Some research has also examined the influence of pragmatic information on inference (Spellman & Holyoak, 1996). These studies demonstrated that goals active when processing an analogy can influence what information people place in correspondence when making a mapping, and can also influence what facts from the base domain are drawn as inferences.

In the present studies, we examine the relative strengths of systematicity and pragmatics

Base Domain

Biology Department
Great teachers *causing*
Students to be motivated to learn
Faculty to get external offers and leave

Political Science Department
Faculty argue *causing*
Faculty to be inaccessible
Department to split into two departments

Target Domain

Music Department
Great teachers

Faculty Argue

Figure 1. Illustration of part of the base and target domains for the experiments. The actual materials were paragraph descriptions of departments. In the studies, there were descriptions of four departments in the base, and two departments in the target. Half of the causal consequents in each department were relevant to the student context, and half were relevant to the financial office context.

as constraints on analogical inference. For this purpose, we adapted materials used by Markman (1997). The design of the study is shown in Figure 1. Participants were given descriptions of departments in a college. The description of each department contained a causal antecedent (e.g., The faculty in the biology department are great teachers) and two consequents following from that antecedent (e.g., students [in biology] are motivated to learn, and faculty [in biology] get external offers and leave the university).

The descriptions of the departments were constructed so that half of the causal consequents in the base domain were most relevant to students at the university, and half were most relevant to financial officers. All materials were pretested to ensure that the causal consequents were primarily related to only one of the contexts.

In Experiment 1, half of the subjects are told at the beginning of the study that they are playing the role of a student at one university who is about to transfer to a second university. The other half of the subjects are told that they are playing the role of a financial officer at one university who is about to take a job at a second university. After reading the descriptions of the departments in the old university (the base domain), they are given descriptions of two departments in the new university (the target domain). Subjects are told that they do not know too much about the new university yet, because they are just arriving, and they are asked to make predictions about what to expect at the new school based on what they know about the old school.

Based on previous research, we expect that the inferences people generate will generally reflect shared system inferences rather than nonshared system inferences. Further, people should tend to infer information that is relevant to them. That is, subjects in the student condition should tend to infer student-relevant information, while subjects in the financial officer condition should generally infer financial officer-relevant information.

A key question involves which constraint will be more important in inference. One possibility is that people will make primarily shared system inferences, but that within the shared system inferences made, there will be more student-relevant facts inferred by people given the student cover story, and more financial officer-relevant facts inferred by people given the financial officer cover story. A second possibility is that pragmatics is most central. On this view, people will focus primarily on pragmatically relevant information regardless of whether it is a shared system fact or a nonshared system fact.

EXPERIMENT 1

Method

Subjects

Subjects in this experiment were 48 undergraduates at Columbia University (24/condition), who were paid to participate.

Design

The main dependent variable in this study is the number of inferences made. The inferences made can be scored as Shared or Nonshared system inferences. Half of the facts in the base domain are Student-relevant, and half are Financial Officer-relevant. The independent variable in this study is Cover story, which has two levels (Student and Financial Officer).

Materials and Procedure

The experimental materials were placed in booklets. The booklets began with instructions that described the cover stories. In the student condition, the subject was told that they were a student at one university (Gordmont University) and that they were transferring to a second university (Fallsburg University). In the financial officer condition, subjects were told that they were a financial officer who worked at Gordmont University, and they were taking a new job at Fallsburg University.

After the cover stories were the descriptions of four departments at the first college. As summarized in Figure 1, each department consisted of paragraphs describing a fact that served as a causal antecedent. This antecedent caused two consequents. One consequent was pretested to be relevant primarily to students, and the other was pretested to be relevant primarily to financial officers. The consequents were judged by the authors to be plausible consequences of the causal antecedent.

After reading the descriptions of the base domain, subjects were given a quiz. Markman (1997) used a similar quiz to ensure that subjects actually read the information about the base domain carefully. The quiz had one question relevant to each causal consequent.

Following the quiz, subjects were shown the descriptions of two departments at the new college. These descriptions were shorter, and contained information about possible causal antecedents, without any information about what occurred as the result of these antecedents.

After reading about the departments at the new school, people were asked to use their experience at the old school to make predictions about what might happen at the new school. Subjects were encouraged to make as many predictions as they wanted.[1] Subjects were given only one booklet for this class, and so they could go back and look at the base domain when making inferences.

Following the inference task, people were asked to say which departments in the old school corresponded to each department in the new school. No specific predictions are made about performance in this mapping task, and it will not be discussed further in this paper.

Results

Inferences were coded as shared system facts, nonshared system facts or other. To be scored as a shared system or nonshared system inference, the subject had to mention a particular causal antecedent and a fact that followed from it. Shared system inferences were those inferred items that were causal consequents from a shared causal antecedent. For example, inferring that faculty will get external job offers and leave given that faculty in the department were good teachers would be a shared system inference. Nonshared system inferences were those for which the inferred causal consequent was not connected to a matching antecedent from the base. For example, inferring that the faculty in a department argue, which will cause them to get outside offers and leave would be a nonshared system inference. All other inferences were scored as other. In the interest of space, inferences scored as other will not be discussed further in this paper.

Each inference was also scored as student relevant or financial officer relevant These determinations were based on how a fact was classified based on the pretests described above.

The data were analyzed in a 2 (shared vs. nonshared system inference) x 2 (student-relevant vs. financial officer-relevant) x 2 (Cover story) mixed model ANOVA. As expected, people made more shared system inferences (M=2.31) than nonshared system inferences (M=0.54), $F(1,46)=58.45$, $p<.001$. The only other reliable effect was an expected interaction between Cover story and Relevance of fact, $F(1,46)=5.25$, $p<.05$. This interaction reflects that subjects given the student cover story made inferences of significantly more student-relevant facts (M=1.75) than financial officer-relevant facts (M=1.08), $t(23)=2.56$, $p<.05$ (Bonferroni). In contrast, students given the financial officer cover story made inference of more financial officer-relevant facts (M=1.50) than student-relevant facts (M=1.37), although this difference was not significant, $t(23)=0.55$, $p>.10$ (Bonferroni).

---

[1] Unlike the studies by Markman (1997), the question in the inference task was open-ended. In previous studies, subjects were asked to make predictions about what would happen given particular facts about the new school. This task may have focused people on specific facts, and inflated the importance of shared system facts connected to those causal antecedents. Thus, these previous studies may have overestimated the importance of shared system facts in inference. The open-ended question does not focus people on particular causal antecedents, and so does not lead to the same potential bias. Because the data from the present studies are similar to those of previous studies, it is unlikely that the phrasing of the question in those studies inflated the importance of shared system facts.

Discussion

These data demonstrate both effects of systematicity and pragmatic relevance on analogical inference. First, replicating previous research, people were far more likely to infer shared system facts than to infer nonshared system facts. In addition, they were more likely to give information that was relevant to their cover story than to give information not relevant to their cover story.

These data further suggest that systematicity is a stronger constraint on inference than is pragmatics. In particular, there were many shared system facts inferred that were not relevant to a subject's cover story. In contrast, there were few nonshared system facts inferred overall. Thus, people appear to filter the inferences they make first by focusing on shared system facts. Once the shared system facts have been found, people can then focus more selectively on those relevant to their goals.

One possible explanation for why the influence of structure appeared stronger than the influence of pragmatics is that people were able to look back at the base domain when making inferences. This explanation assumes that one important role of pragmatic goals is to focus people on information that is likely to be relevant when faced with a heavy memory load. Because people were able to look back at the base and target domains in Experiment 1, there was no significant memory load. Thus, pragmatic information may have been less useful than it would be if the base domains were in memory.

To test this possibility, we repeated Experiment 1, except that there were two booklets. One contained the base domain and the quiz. The other contained the target domain and the inference and mapping tasks. At the beginning of the study, subjects were given the base domain and the quiz. After completing the quiz, the first booklet was taken away, and the second booklet with the target domain and the inference and mapping tasks was given. Thus, subjects had to recall information about the base domain from memory. If pragmatic information has its influence primarily on memory, then the effects of pragmatics relative to those of structure should be stronger in Experiment 2 than they were in Experiment 1. Otherwise, the data are expected to look much like those of Experiment 1.

EXPERIMENT 2

Method

Subjects

Subjects in this study were 48 members of the Columbia University community who were paid for their participation.

Materials, Procedure, and Design

The materials, procedure, and design of Experiment 2 were identical to those of Experiment 1 with the following change. The booklets generated in Experiment 1 were split into two parts. The first part contained only the instructions with the cover story, the description of the base domain, and the quiz. The second part contained the description of the target domain, the inference task and the mapping task. After completing the quiz in the first part, subjects had to turn in the first booklet in order to receive the second and to complete the experiment.

Results

Once again, the data were scored as shared system facts and nonshared system facts. in addition, the information was marked as student-relevant or financial officer-relevant. Again, the data were analyzed with a 2 (shared vs. nonshared system inference) x 2 (student-relevant vs. financial officer-relevant) x 2 (Cover story) mixed model ANOVA.

The results of this study are quite similar to those of Experiment 1. Once again, people made more shared system inferences (M=2.29) than nonshared system inferences (M=0.58), $F(1,46)=42.23$, $p<.001$. There was also a reliable interaction between Cover story and Relevance of fact $F(1,46)=4.10$, $p<.05$. As before, this ANOVA reflects that people given the student cover story made more student-relevant inferences (*M*=1.46) than financial officer-relevant inferences (*M*=1.25) and people in the financial office condition made more financial officer-relevant inferences (*M*=1.75) than student-relevant inferences (*M*=1.29). Neither of these simple effects was reliable, however.

Finally, there was a marginally significant interaction between Inference type and Relevance, $F(1,46)=2.91$, $.05<p<.10$. This interaction reflects that, collapsing across cover stories, there was a tendency for people to make fewer shared system inferences of student-relevant facts (*M*=1.04) than of financial officer-relevant facts (*M*=1.25), but more nonshared system inferences of student-relevant facts (*M*=0.33) than of financial officer-relevant facts (*M*=0.25). Neither of these simple effects is reliable.

Discussion

The results of Experiment 2 provide additional evidence that systematicity is a more powerful constraint on analogical inference than is pragmatics. As in Experiment 2, people made far more shared system inferences than nonshared system inferences. There was a tendency for people to make inferences of facts related to their pragmatic goals, but this tendency was small relative to the influence of systematicity.

Experiment 2 extends the findings of Experiment 1, because people could not look back at the base domain when making inferences in this study. We speculated that being able to look back at the base domains might have decreased the influence of pragmatic goals. In contrast to this speculation, having to access the base domain from memory did not make the influence of pragmatic goals on inference stronger.

While pragmatic goals have a weaker influence on analogical inference than does systematicity, they have still had a reliable influence on inferences in two studies. Thus, it is worth considering where these goals have their influence. Spellman and Holyoak (1996) contrasted two possible influences of pragmatics. One possibility was that pragmatics influenced pre-mapping representational processes. On this view, information about domains is filtered by pragmatic goals, and only goal-relevant information is stored. Alternatively, pragmatic goals might have their influence during the mapping and inference phases.

To test this possibility, Spellman and Holyoak (1996) varied when in the experiment people were given pragmatic information. It was either given prior to the presentation of the base domain (in which case it could have some influence on what was stored) or after the presentation of the base domain (in which case, it could not have influenced what was stored). Regardless of when pragmatic information was presented, an influence of pragmatic goals was found, leading Spellman and Holyoak to conclude that pragmatics has its influence after the domains are represented.

We performed a similar study with our materials. As in Experiment 2, the base and target domains were presented in separate booklets. The subjects in this study were given the cover story after taking the quiz and receiving the second booklet. This group of subjects had already seen the base domain, and so the pragmatic information could not influence what was learned about it.

If, as Spellman and Holyoak (1996) suggested, pragmatic information has its influence after the representation process is completed, then we should obtain the same results in Experiment 3 that were observed in the first two studies. In contrast, if pragmatic information has its influence during the construction of representations, then the influence of the cover story should be eliminated in Experiment 3.

EXPERIMENT 3

Method

Subjects

Subjects in this experiment were 48 members of the Columbia University community (24 per condition) who were paid for their participation.

Materials, Procedure, and Design

This experiment was identical to Experiment 2, except that the cover story was presented to subjects after completing the quiz, and before the read about the target domain.

Results

Once again, the inferences were scored as shared system facts, nonshared system facts or other facts. The shared and nonshared system facts were further scored as student-relevant or financial officer-relevant. The data were analyzed with a 2 (shared vs. nonshared system inference) x 2 (student-relevant vs. financial officer-relevant) x 2 (Cover story) mixed model ANOVA.

As in Experiments 1 and 2, people made more shared system inferences overall (M=2.15) than nonshared system inferences (M=0.75), $F(1,46)=63.77, p<.001$. In addition, as in Experiments 1 and 2, there was a significant interaction between Cover story and Relevance of fact, $F(1,46)=11.12$, $p<.01$. This interaction reflects that subjects given the student cover story made more student-relevant inferences (M=1.96) than financial officer-relevant inferences (M=0.82), $t(23)=3.87$, $p<.01$ (Bonferroni). In contrast, subjects given the financial officer cover story made more financial officer-relevant inferences (M=1.54) than student-relevant inferences (M=1.46), although this difference was not significant, $t(23)=0.39$, $p>.10$ (Bonferroni).

In addition to these expected effects, there were also two unexpected effects. There was a main effect of Relevance of fact, $F(1,46)=8.27, p<.01$. This interaction reflects that overall there were more student-relevant inferences (M=1.71) than financial officer relevant inferences (M=1.19). Finally, there was a reliable interaction between Cover story and Shared vs. nonshared system, $F(1,46)=4.11, p<.05$. This interaction reflects that people tended to make slightly more shared system inferences when given the financial officer cover story (M=2.38) than when given the student cover story (M=1.92), but to make slightly more nonshared system inferences when given the student cover story (M=0.88) than when given the financial officer cover story (M=0.63). Neither of these simple effects is significant.

Discussion

The results of Experiment 3 are parallel to those of Experiments 1 and 2. Once again, there was a strong tendency for people to infer shared system facts rather than nonshared system facts. In addition, people were also more likely to infer facts relevant to the cover story given rather than facts not relevant to the cover story.

The strong influence of pragmatics in this experiment suggests that pragmatic information has its influence after the representations of the domains have been formed. That is, people could not filter out information about the base domain using their goals, because these goals were not presented until after the base domains were encoded. This pattern of data is sensible, because people often cannot know their goals in advance. Encoding as much information as possible is advantageous, because it allows information relevant to an unforeseen goal to influence cognitive processing.

An unexpected finding in this study was that people inferred more student-relevant facts overall than financial officer-relevant facts. This finding may reflect an influence of background knowledge on memory. In this study, people were not given the cover story until after they read about the base domain. Thus, they had to bring their own experience to bear when interpreting the base domain. Because all of the participants in this study were students, it is likely that they found the student-relevant information more salient or more comprehensible than the financial officer-relevant information. This suggestion is compatible with a variety of studies demonstrating that memory for new information in familiar domains is better than memory for new information in unfamiliar domains (Bransford & Johnson, 1972, 1973).

GENERAL DISCUSSION

Analogical inference is a powerful way of extending one domain based on its similarity to another. Because much of the information about a base domain is unlikely to be true in the target domain, it is necessary to constrain the inference process. The best constraints are those that focus people on the information in the base domain that is most likely to be true about the target.

Two constraints on analogical inference examined here were systematicity and pragmatics. Strong support for the influence of systematicity was obtained in these studies, as subjects inferred far more shared system facts than nonshared system facts. Support for pragmatics was also obtained, as people were generally more likely to infer facts relevant to their cover story. This influence of pragmatics was evident even in Experiment 3 where the goal was not provided until after the base domain was read. This finding suggests that pragmatics does not filter out information during encoding, but rather works during the mapping or inference stage. Further research will have to pinpoint the locus of the effects of pragmatics.

While both systematicity and pragmatics had an influence on analogical inference, systematicity was a much stronger constraint in these studies. People generally inferred causal consequents that were related to matching causal antecedents. Having used systematicity to constrain the set of possible inferences, people were then somewhat more likely to infer information relevant to their cover story. However, in all conditions, there were still many inferences of information not relevant to the cover story. It is possible that the effects of pragmatics would be stronger if the consequences for failing to achieve the goal were more dire. In the present experiments, people were simply given a cover story, but were not rewarded selectively for inferences relevant to their cover story, or penalized for inferences not relevant to their cover story. Nonetheless, the present results strongly suggest that systematicity is a more powerful constraint on inference than is pragmatic relevance.

Implications for Computational Models

There are a number of comprehensive models of analogical reasoning, and all of them have mechanisms for generating analogical inferences.[2] In this discussion, we focus on three prominent models: SME (Falkenhainer, Forbus, & Gentner, 1989), LISA (Hummel & Holyoak, 1997), and IAM (Keane, Ledgeway, & Duff, 1994).[3] This discussion will assume a basic familiarity with these models.

The SME model assumes that candidate inferences involve carrying over facts from the base domain that are connected to matching systems. Thus, SME is consistent with the observed use of systematicity to constrain analogical matches. SME has been extended to incorporate pragmatic information as well (Forbus & Oblinger, 1990). This extension marks pragmatically relevant representational elements, and then attempts to use the goal relevant information in the preferred mapping, and in the candidate inferences generated. This use of pragmatics is consistent with the idea that systematicity is a stronger constraint on analogical inference than pragmatics.

There are two ways in which SME has difficulty explaining the present data. First, the implementation of pragmatic marking in SME is too strong. In the data, pragmatic information appears to provide a small increase in the salience of facts relevant to the goal. In contrast, SME will carry over every marked fact that is structurally consistent with the match between base and target. Thus, in order to allow SME (with pragmatic marking) to account for the present data, some mechanism must be established to determined how nodes are given pragmatic marking. This account would have to assume that not all representational elements that are goal-relevant get marked, or that not all relevant facts are posited as candidate inferences.

---

[2] There are also many specialized models that do not incorporate analogical inference mechanisms. For example, Halford et al.'s STAR model uses tensor products in a connectionist model to do A:B::C:D analogies (Halford, Wilson, Guo, Wiles, & Stewart, 1994). Candidate inferences are not needed to solve this type of analogy problem.

[3] Holyoak and Thagard's (1989ACME is not considered here. Its candidate inference mechanism is not constrained by systematicity, and has difficulty making inferences when there are potential many-to-one matches (Markman, 1997).

Second, some information that is not goal-relevant is also inferred by subjects in the present studies. Thus, the pragmatic marking account must also explain why some (but not all) non-goal-relevant facts are inferred.

The IAM model generates candidate inferences by completing partially matching systems. This assumption constrains IAM to infer only shared system facts. Pragmatic information influences IAM by determining which predicates are used for the match between base and target. Goal-relevant predicates are more likely than non-goal-relevant predicates to be selected at the early stages of the match process to be parts of the correspondence. In the end, however, IAM generates a match that includes both relevant and irrelevant matches in a situation like the one in the present studies, because both the relevant and irrelevant information can be incorporated into a structurally consistent match. Thus, IAM cannot explain why goal-relevant inferences were more common than non-goal-relevant inferences.

Finally, it is not clear what LISA predicts for this task. Comprehensive tests of the candidate inference mechanism in LISA model have not yet been published, but Hummel (personal communication) suggests that LISA exhibits a preference for shared system facts over non-shared system facts in analogy. There are a number of ways that pragmatic information could be incorporated into LISA. Relational bindings are represented in this model by having nodes that correspond to predicates, relational roles, and arguments to those relations fire in phase with one another, and out of phase with nodes representing other relational bindings. Pragmatically relevant bindings can be fired more often than pragmatically irrelevant bindings. A mechanism like this would help ensure that pragmatically relevant information is incorporated in the mapping that is generated. It is possible that this mechanism would also lead to more goal-relevant inferences than non-goal-relevant inferences. At this time, however, it is not possible to make any firm predictions.

CONCLUSIONS

The three experiments in this paper demonstrate that systematicity and pragmatics are important constraints on analogical inference. Shared system facts are more frequently inferred than nonshared system facts. Likewise, goal-relevant facts are more frequently inferred than non-goal-relevant facts. Further, systematicity appears to be a more powerful constraint on mapping than is pragmatics.

Currently, none of the comprehensive computational models accounts for all of the data. All of the models have mechanisms for implementing both systematicity and pragmatics. However, SME cannot account for why some goal-relevant facts are not inferred, while some non-goal-relevant facts are inferred. IAM cannot account for why there is a difference in the number of goal-relevant and non-goal-relevant facts inferred. Finally, LISA exhibits a preference for shared system facts, but its inference mechanism has not been specified to the point where it can make specific predictions about the role of pragmatic information in inference. Further research on these computational models will have to address these shortcomings.

AUTHOR IDENTIFICATION NOTES

This work was supported by NSF CAREER award SBR-95-10924 given to the first author. The authors would like to thank Dedre Gentner, Keith Holyoak, John Hummel, Julia Kalmanson, Tomislav Pavlicic, and Takashi Yamauchi for helpful comments during the evolution of this project.

REFERENCES

Andersen, S. M., & Cole, S. W. (1990). "Do I know you?": The role of significant others in general social perception. *Jour-*

nal of Personality and Social Psychology, 59(3), 384-399.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.

Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In W. G. Chase (Eds.), *Visual Information Processing* (pp. 383-438). New York: Academic Press.

Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science, 15*, 89-132.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41*(1), 1-63.

Forbus, K. D., & Oblinger, D. (1990). Making SME greedy and pragmatic. In *The Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*. Boston, MA: Lawrence Erlbaum Associates.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 199-241). New York: Cambridge University Press.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306-355.

Halford, G. S., Wilson, W. H., Guo, J., Wiles, J., & Stewart, J. E. M. (1994). Connectionist implications for processing capacity limitations in analogies. In K. J. Holyoak & J. Barnden (Eds.), *Advances in Connectionist and Neural Computation Theory, Vol. 2: Analogical Connections* (pp. 363-415). Norwood, NJ: Ablex.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*(3), 295-355.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*(3), 427-466.

Keane, M. T., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science, 18*, 387-438.

Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science, 21*(4), 373-418.

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier examples. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*(3), 456-468.

Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology, 31*, 307-346

PRAGMATIC EFFECTS ON SPEED OF ANALOGICAL PROBLEM SOLVING

Bruce D. Burns
Institut für Psychologie, Universität Potsdam
14415 Potsdam, Germany
(burns@rz.uni-potsdam.de)

ABSTRACT

Gentner & Holyoak (1997) pointed out that there has been convergence between theories of analogy. However, the role of pragmatics in analogy appears to still divide theories. The effect of pragmatics on the speed of analogical problem solving was investigated using highly simplified chess problems. The pragmatic factor of goals was manipulated by instructing participants to make an attacking or defensive move. Participants received training problems, followed by a set of testing problems which were solvable by analogical transfer from a training problem. It was found that presenting the same goal at test as was given in training for a maneuver led to faster solutions, but the effect of piece similarity (which determined structural similarity) interacted with goal similarity. Piece similarity helped when the pragmatics were consistent, but when the pragmatics were inconsistent, other forms of similarity had no effect. This supports theories in which pragmatics acts as a strong filter for analogies, rather than an attenuated filter.

PRAGMATICS AND ANALOGICAL PROBLEM SOLVING

Gentner and Holyoak (1997) pointed out that a consensus as to the nature of analogical reasoning has emerged. However, in the companion pieces introduced by Gentner and Holyoak (i.e., Gentner & Markman, 1997; Holyoak & Thagard, 1997), a stark difference is apparent: pragmatics is emphasized by Holyoak and Thagard, but ignored by Gentner and Markman. The role of pragmatics appears to remain a point of dispute between theories of analogical reasoning.

According to Holyoak and Thagard (1989), the pragmatics of an analogy are the goals and purpose of the analogist. The context may provide such pragmatics, or they may be bought by the analogist to the situation, either way they will influence what analogies may be formed. It is not disputed that pragmatics are important for analogical reasoning, but how is. Pragmatics were implemented in Holyoak and Thagard's ACME computer model as providing emphasis for important mappings or elements of an analog. In contrast to pragmatics affecting the process of analogical mapping, Gentner (1989) argued that pragmatics could have an influence before processing, by changing the representation of the analogs; alternatively, pragmatics could have an influence after processing, by causing the rejection of analogies; but pragmatics have no independent effect during processing. In the implementation of Gentner's ideas in SME (see Falkenhainer, Forbus & Gentner, 1989) analogical processes are driven by structural and semantic factors alone.

Spellman and Holyoak (1996) found evidence that pragmatics influenced the process of analogical mapping by showing that *process* goals (i.e., the goals of the reasoner rather than those contained within the analogs) influenced the mappings people made. In particular, pragmatics did not filter out all goal-irrelevant information, as it would if pragmatics selected the relevant parts of the source and target as input to the mapping process. Rather than being a strong filter, as attention was in Broadbent's (1958) selective attention model, pragmatics instead was an attenuated filter, as in Treisman's (1964) alternative to Broadbent's model. Such an attenuated filter does not completely block out the information it filters.

Therefore, Spellman and Holyoak (1996) derived two testable hypotheses, the *filter hypothesis* and the *filter-attenuation hypothesis*. Fundamentally Spellman and Holyoak's argument appears to make predictions about how pragmatics would interact with other factors such as semantic similarity and structural consistency. If pragmatics are a strong filter, then other factors should have no influence on analogical success when the pragmatics are wrong. If pragmatics are an attenuated filter, then other factors should influence success even when the goal is wrong. Therefore, the filter hypothesis (here referred to as the *strong filter hypothesis*) could be contrasted with the attenuated filter hypothesis by crossing pragmatics with other factors experimentally.

If the argument over the role of pragmatics is over its effect on processes, then response times may be a particularly appropriate dependent measure. Many investigations of cognitive processes have used response times (see Posner, 1986), yet response time has rarely been used to investigate analogical processes. Klein (1986) argued that speed should be an advantage of analogical thinking, but did not directly test this idea. Thus using response time as a dependent measure allowed the validity of speed as measure of analogical problem solving to be examined, and opened up the possibility of gaining insight into analogical problem solving as a process.

CRITERIA FOR I NVESTIGATING PRAGMATICS

It is inherently difficult to explore what happens during a cognitive process, and exploring the role of pragmatics in analogical processes raises a unique set of problems. Spellman and Holyoak (1996) proposed three criteria that need to be fulfilled.

1. The pragmatic constraints must not be reducible to other general constraints. If goals simply form parts of the structure of an analog, then their influence could be explained as a special case of the influence of structural and/or semantic constraints. In that case pragmatics would be just like any other shared representational component. To clarify this issue, Spellman and Holyoak (1996) distinguished between *static* and *processing* goals. For example, in mapping the 1991 Persian Gulf War to World War II, Hitler's goal of taking over Europe could be mapped to Saddam Hussein's assumed goal of taking over the Persian Gulf. This would be a mapping of static goals internal to the analogs. Spellman and Holyoak argue that the Bush administration promoted the mapping between the Persian Gulf crisis to World War II in order to achieve an external goal: military intervention by the United States in the Persian Gulf.

However, the distinction between a processing goal and static goal is problematic. One problem acknowledged by Spellman and Holyoak (1996), is that it is difficult to rule out that a processing goal is immediately converted into a static goal once it is given. A further problem is that it may not be clear which goals are internal or external to an analog. For example, it could be argued that the Bush's military intervention goal in 1991 was a static goals internal to the analogs. The World War II analog could already have a military intervention goal embedded within it, as a result of the perceived failure of appeasement before World War II. Thus it could be argued that the pre-World War II analog was retrieved by the Bush administration because people represented it with a static goal that mapped to Bush's own static goal for the Gulf War crisis, that the United States should intervene.

2. The pragmatic effects should not be attributable to post-mapping processes. When analogies are used to solve problems, as they have been in many studies of analogy, then processes after the mapping process are required. Mappings which violate the goals will be rejected. To avoid this problem, Spellman and Holyoak (1996) focus on the actual mappings people make rather than what they do with that mapping. However, even mapping tasks are vulnerable to post-mapping processes, because not all mappings can be recorded simultaneously.

3. The pragmatic effects should not be attributable to pre-mapping processes. Pragmatics may change the representation of analogs before the mapping process begins. Establishing that mapping depends on goals given after initial representation was the major purpose of the experiments by Spellman and Holyoak (1996). By finding empirical support for the filter-attenuation hypothesis, they found support for the claim that pragmatics affects the process of analogical mapping.

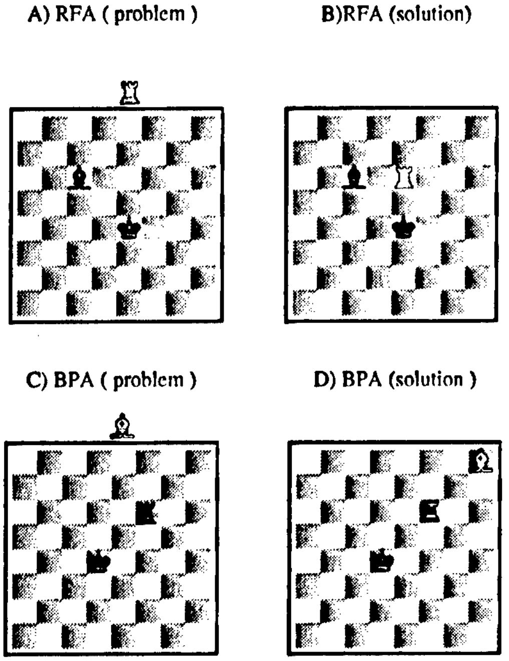
Meeting these criteria. In this experiment an analogical problem solving task was used to try and meet the above criteria. This was partly because speed was used as the measure of analogical reasoning rather than success, in which case a problem solving task was more appropriate than a mapping task (it is more likely to have a clear and definite end). Problem solving also has an advantage in that it has a very clear processing goal: solve the problem by achieving the specified goal. This processing goal has a definite and consistent focus on mapping the goals of the source analog. Therefore, rather than manipulating processing goals, an alternative way to the investigate the effects of processing goals is to maintain a consistent processing goal, but manipulate the nature of the arguably static goals that it focuses on. This manipulation would allow the same hypotheses to be tested as when the processing goals are manipulated.

For minimizing the influence of post-mapping processes, problem solving tasks for which it is obvious if the solution is correct, could be particularly appropriate. Clear solutions, that require no modification, are less likely to invoke post-mapping processes.

THE TASK

To determine the effects on analogical problem solving of pragmatics, required a problem solving task with goals that were easy to manipulate without affecting other factors. Such a task is chess which contains a clear distinction between the goals of attack and defend. Just as importantly, in chess the exact same configuration of chess pieces can be approached by a player as a position in which an attack should be launched (i.e., an attempt should be made to capture opposing pieces or to gain a more favorable position), or as one to be defended (i.e., your own pieces should be protected, or your position should not be allowed to deteriorate). Thus chess has clear pragmatics that can be manipulated independent of the structure of a position (i.e., the relationships between pieces) and its semantic components (i.e., the actual pieces themselves). So the problems consisted of highly simplified chess positions. For each problem, participants were presented with a chess board on which were placed two defender chess pieces. One attacker piece was presented off the board, waiting to be placed on the board (see Figures 1a and 1c for examples of exactly of what participants saw).

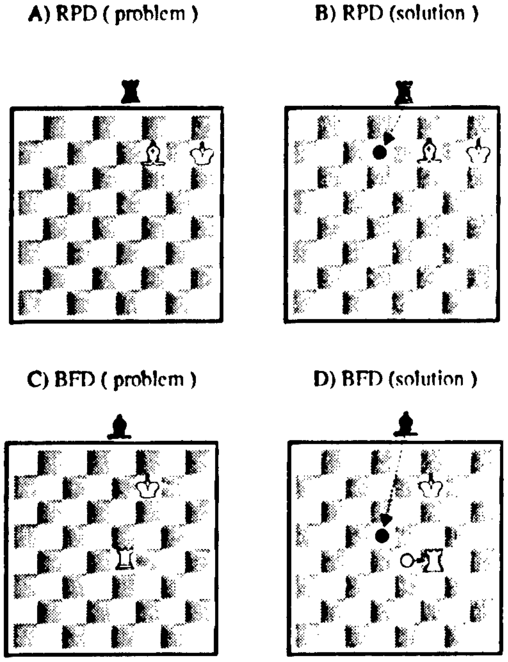
When the goal was attack, participants solved the problem by placing the attacker piece onto the board so as to guarantee that the attacker piece would be able to capture one of the defender pieces on its next move (after one of the defender pieces had had the opportunity to make one move, just like in normal chess). Example solutions for the problems in Figures 1a and 1c are shown in Figures 1b and 1d respectively. Identical positions were given when the goal was defend, but the problem task was the opposite: the participant had to avoid the capture of a defender. The participant legally moved one of the two defender pieces in anticipation of the attacker piece being placed onto the board (e.g., Figures 2a and 2c). This attacker piece would be anticipated to make an attacking maneuver, exactly like the one that the paricipants would have made if they had the attack goal. For example, Figures 2b and 2d show solutions to the problems shown in Figures 2a and 2c respectively. The defend goal thus incorporated the attack goal.



A) RFA (problem) B)RFA (solution)

C) BPA (problem) D) BPA (solution)

Figure 1. Examples of attack goal problems: (a) is a rook fork and (c) a bishop pin, each with the attacker piece (white) shown above the board and two defenders pieces (black) on the board. Solutions to these problems are indicated in (b) and (d) respectively, which show successful placement of the attacker.



A) RPD (problem) B) RPD (solution)

C) BFD (problem) D) BFD (solution)

Figure 2. Examples of defend goal problems: (a) is a rook pin and (c) a bishop fork, each with the attacker (black) shown above the board and two defenders (white) on the board. In (b) and (d) are solutions to these problems, each a successful move of a defender (an open circle and line show where the piece moved from). Also shown is where the attacker was threatening to go (indicated by the solid circle and dashed line).

To help participants solve these problems, they were trained on two simple chess tactics known as *pins* and *forks*. Figure 1b illustrates a fork solution: an attacker piece is placed so as to simultaneously attack two defender pieces. Because only one defender piece can be moved at a time, only one defender will be able to escape the attack, leaving the other to be captured. Figure 1d illustrates a pin solution: the attacker was placed such that only one defender was directly threatened, but if this first defender moved away then the defender behind it could be captured. Hence, a capture was guaranteed.

When the goal was defend the participant had to anticipate that the opponent was about to place the attacker piece onto a square from which it could execute a pin or a fork. The participant must move one of the defender pieces so that no matter where the attacker piece was placed, it could not execute a pin or fork. An example of a problem requiring defense against a rook pin is illustrated in Figure 2a, and a successful solution is shown in Figure 2b (which also illustrates the threat). Figure 2c illustrates a defend goal when a fork by a bishop was threatened, and Figure 2d is a solution.

Defense and attack are closely related in these problems as identical configurations of pieces could be used for both goals. Knowing how to successfully attack should help with achieving the defend goal, as the specific attack solution was what had to be defended against. Similarly, knowing how to defend a position implies knowing how to attack that same position.

General methodology

Participants received training on achieving both attack and defend goals, and training on how to execute pins and forks. They were then tested by being presented with problems that required a pin or fork solution, but varied in whether they had the same goal or used the same attacker piece as did the participant's training for pins.

There were eight basic training positions possible, each a combinations of the three two-level factors of bishop/rook attacker, pin/fork solution, and attack/defend goal. These positions will be referred to by combinations of their initials: a *bishop-pin-attack* problem will be referred to as BPA, a *bishop-pin-defense* as BPD, a *bishop-fork-attack* as BFA, a *bishop-fork-defense* as BFD, a *rook-pin-attack* as RPA, a *rook-pin-defense* as RPD, a *rook-fork-attack* as RFA, and a *rook-fork-defense* as RFD. Participants were trained on one of the four pin problems.

The two independent variables of interest were *goal change* (i.e., pragmatics) and *piece change* (i.e., structure). Relative to a player's own training on a certain solution type (i.e., pin or fork), all test positions could be considered either the present or absence of change along either or both of these dimensions. A goal change was defined as a problem with the opposite goal to the pin training problem. A piece change was defined as changing the attacker piece, from a rook to a bishop. Changing the attacker required changing the relationship between the defender pieces and required thinking about the problem in a different way because of the different ways that rooks and bishops move, thus it changed the structure of the problem.

Crossing the two variables yielded four different types of problems: *no-change*, problems that used the same goal and attacker piece as a participant's pin training problem; *change-piece*, problems with the same goal, but different attacker piece; *change-goal*, same attacker piece, different goal; and *change-both*, problems with a different goal and attacker piece from the training problem. For example, a BPA test problem was a change-goal problem if the participant received BPD training, but a change-piece problem if the participant received RPA training. Thus, because different participants received different training problems, every specific test problem was classified into one of these four types, depending on a participant's specific pin training. Thus, the design was completely within-subject and the effects of differences in the difficulty of specific problems was eliminated by having equal numbers of participants experience each type of training.

In order to increase the number of testing problems and to examine the effects of surface changes to the problems, a third type of change was applied to problems independent of the piece and goal changes: *Surface transforms,* which were transformations of the training problems involving changing the placement or nature of defender pieces.

Predictions

Both the strong filter and attenuated-filter hypotheses would predict that a pin problem with the same goal as the pin training problem should be solved faster than when the problem had a different goal. Structural changes from the pin training problem should also be responded to slower. However, how surface and structural changes interact with goal changes was the critical question with regard to testing the predictions of the strong- and attenuated-filter hypotheses.

The strong filter hypothesis would appear to predict that the goal and structure changes should interact, such that when a problem had the same goal as the training problem requiring the same solution, then having similar structure should further speed responding. In contrast, when the goal is different, the strong filter should render other factors irrelevant. Thus problem solvers with the wrong goal would not be helped by similar structure or surface features, because they would be searching the wrong part of memory (see Schank, 1982) or because pragmatics had an effect outside the process of analogical mapping (see Gentner, 1989). Therefore, when the problem has the wrong goal, neither structure nor surface similarity should affect the speed of solutions.

The filter-attenuation hypothesis, suggested by Holyoak and Thagard's (1989) multiconstraint theory, could be consistent with either the presence or absence of an interaction between goal and structure manipulation. The critical prediction of this hypothesis was that other factors should continue to have an effect even when the goals were wrong. If pragmatics are part of the process of analogical mapping, then it would be expected that structural features would continue to influence problem solving even when the goal was wrong. Therefore, the strong filter and filter-attenuation hypotheses make contrasting predictions for the effects of structure, when the goals of the source and target analogs do not match.

AN EMPIRICAL TEST

Method

Participants. A total of 108 participants (87 male and 21 female), with 27 in each pin training group were drawn from the introductory psychology participant pool at University of California, Los Angeles.

Apparatus. An Apollo series 4000 workstation with a 19 in. color monitor and a three-button mouse was utilized. A program developed by the author controlled the experiment and response times were measured with an accuracy of one second.

Materials. The pin training problems were similar to those shown in Figures 1c and 2a, except that both attacking and defending versions of either could be given. This yielded four training problems. An additional change was that the non-king defender was always a knight, rather than varying with the attacker piece. The same type of fork training problem was given to all participants. This problem used a knight as the attacker and the defenders were a king and a queen. Thus, the fork training problems had minimum similarity with any of the pin problems.

Forty-two testing problems were given, all of which were solvable with a pin or a fork. Most of these were based on the four pin training problems specified, as well as on the four possible fork training problems not given in this experiment. However, one of four transforms were applied: the *identity* transform, was identical to a training problem; the *rotate* transform, rotated a training problem by 90°; the *defender* transform, changed the non-king defender into the opposite type of piece to the attacker piece (i.e., if a rook was the attacker, then the defender was a bishop, and vice verse); the *reconfigure* transform increased the distance between the defenders and changed the non-king defender into the same type of piece as the attacker piece. In addition to these problems, a set of problems to which either a fork or pin solution could be applied were given. These ambiguous problems will not be discussed here. A single randomly generated order for the 42 problems was created, with attack and defend problems alternating.

Procedure. Participants were given practice trials that tested their knowledge of the moves of chess, and which gave them practice with recognizing the pieces, and with moving them around with the mouse. To teach them about attack and defend goals, they were given knight-fork-attack problems, and then knight-fork-defense problems. For each of these sets of problems, they had to correctly solve three consecutive problems, before they went on to the next stage. Each training problem in a set was the same except that the pieces were shifted to different places on the chess board. The computer showed participants a correct solution if they were incorrect.

The link between attack and defense was made very explicit in the participants' instructions. For defend problems, they were advised to first think of where they themselves would place the attacker, if they had the chance. Then they should move a defender to render that placement harmless.

Two more training sets were then given, the nature of which depended on the training condition of a participant. Participants were given the type of training problem specified by their condition, plus one other set. If they received BPA or RPA training, then they received an extra set of knight-fork-defense problems. However, those given BPD or RPD training received an extra set of knight-fork-attack problems. This equalized the amount of training on attack and defend goals. After the training was successfully completed, participants were given the 42 test problems.

Results

The mean number of errors made by participants was 5.5 (*SD* = 3.14) out of 42. Participants had very high accuracy for attack problems (96% correct), and also high accuracy on defend problems (78% correct).

Response times were skewed so a log transformed was applied to them. Given the high solution rate, the critical dependent measure was log response times for pin problems, ignoring whether a problem was correctly solved. There was no evidence of a speed-accuracy trade-off, as number of errors correlated positively with response time, though not significantly, *r*(108) = .11, *p* = .25. Response times unclassified by change type showed a significant linear trend, *F*(1,104) = 157.30, *p* < .001 (*MSE* = .47), indicating that participants became faster as they completed more problems. There was no effect of training condition on response time, *F*(3,104) = .50 (*MSE* = 2.86), indicating equivalence of the overall effects of training.

Pin problems were then classified by change type (no-change, change-piece, change-goal, change-both) and the mean response times for each of the four change types across the four transformations are presented in Table 1.

no-change	change-piece	change-goal	change-both
2.66 (.41)	2.77 (.54)	2.83 (.45)	2.85 (.45)

Table 1. Mean log response times (SD in parentheses) for each type of problem.

A 4x2x2x2x4 mixed ANOVA was carried out with between-subject factors of training type (four levels) and ambiguous set (two levels), and within-subject factors of goal (same or different), piece (same or different), and transform (four levels). There were main effects of goal, *F*(1,100) = 18.38, *p* < .001 (*MSE* = .35), and piece, *F*(1, 100) = 5.61, *p* = .020 (*MSE* = .30), and an almost significant interaction between goal and piece, *F*(1,100) = 2.99, *p* = .087 (*MSE* = .30). There was also a main effect of transform, *F*(3,300) = 25.24, *p* < .001 (*MSE* = .23), but there were no significant interactions with transform: transform by goal, *F*(3,300) = 1.29, *p* = .28 (*MSE* = .25); transform by piece, *F*(3,300) = .76 (*MSE* = .22); transform by piece by goal, *F*(3,300) = 1.19, *p* = .31 (*MSE* = .26).

The differences between critical groups were tested to determine which of the predicted differences were present. The no-change problems were solved faster than the change-piece problems, *F*(1,100) = 10.04, *p* = .002 (*MSE* = .24). However, change-goal problems did not differ from change-both problems, *F*(1,100) = .17 (*MSE* = .37). Therefore, when the goal did not change, there was a clear effect of piece, but there was no piece effect when the goal was changed. The difference between the change-piece and change-both sets of problems was almost significant, *F*(1,100) = 3.75, *p* = .056 (*MSE* = .34), suggesting that changing the goal had an effect in addition to changing the attacker.

Control comparison. Responses to problems requiring a fork solution allowed a control comparison. If transfer occurred from pin training to fork problems, then fork problems should have been affected by which pin training problem was given. To test this, fork problems were classified as no-change, change-goal, change-piece or change-both, as though they were pin problems. The mean log response times across transforms were: for no change, *M* = 2.86 (*SD* = .41); for change-piece, *M* = 2.90 (*SD* = .49); for change-goal, *M* = 2.93 (*SD* = .45); and for change-both, *M* = 2.93 (*SD* = .49). There was no effect of piece, *F*(1,100) = .70 (*MSE* = .27), or goal, *F*(1,100) = 1.90, *p* = .17 (*MSE* = .41) nor an interaction between these factors, *F*(1,100) = .54 (*MSE* = .25). Surface transform did not interact with anything (all *F*'s < 1.0). Therefore there was no evidence of transfer from pin training problems to fork problems.

Discussion

The experiment found clear effects of goal and piece changes. However, If the goal was the same as that used in the pin training problem, then responding was faster when the same attacker piece was used than when a different attacker piece was used in the problem. Adding a goal change to a piece change slowed responding, but when the goal was changed, there was no difference between piece change conditions. Therefore the results supported the strong filter hypothesis rather than the filter-attenuation hypothesis, as structure was only relevant when the pragmatics matched.

If pragmatics have their effects outside of the mapping process, then the results seem more consistent with explaining the pragmatic effects as due to pre-mapping processes rather than a post-mapping process. If pragmatics had an effect after mapping then a independent main affect of piece would be expected, given that the piece similarity would have an effect before the goal would.

The results do not necessarily disconfirm the claim that goals can affect the process of analogical mapping. As Spellman and Holyoak (1996) pointed out, the strong filter of Gentner (1989) is a special case of the continuum represented by an attenuated filter. Perhaps when the goal is of high importance, the filter may be strong and allow little other information through. Such an argument raises the question of what determines how strong is the filter, otherwise the attenuated-filter hypothesis becomes undisconfirmable.

Processing unsuccessful analogies. The core of the argument over pragmatics concerns how analogies are processed, and the data provide another form of evidence that may address this issue. For many cognitive processes, a key form of evidence has been what happens when they fail. It is known that good analogies can lead to poor inferences when the analogy is inappropriate, but what is the process when an otherwise good analogy fails to lead to any applicable solution? None of the models of analogy explicitly address this issue, nonetheless some intuitions could be derived about what might happen. It would seem that within a serial model, such as Falkenhainer et al's (1989) SME, a goal that leads to an analogy that is inapplicable should result in a slower solution than when the goal does not lead to an inappropriate analogy. Such a goal would be more likely to lead the problem solver down the wrong path, and further down this wrong path, and thus add to the total time to find an appropriate solution (analogically or otherwise). Thus, it would appear that response time should depend on how 'good' the target problem was as an analog to the inapplicable source problem. In contrast, in a parallel model, such as Holyoak and Thagard's (1989) ACME, in which the pragmatics and the solution are all part of the mapping process, then failure would be indicated by failure to converge. When ACME converges, it could be assumed that it will be faster the better the analogy is. However, when it fails to converge it should take the same amount of time to recognize that convergence is not occurring, no matter how good the analogy otherwise appears to be (though this is based on hypothesizing a mechanism in ACME for recognizing failure to converge, something it does not have). Therefore, how good a target is to an inapplicable source analogy should not affect solution time, assuming that it is clear whether a solution can be applied or not.

The fork problem response times provided some empirical data about inappropriate analogies. Participants did not know their pin training would not be applicable to these fork problems, until they tried to map the pin solution to the new problem. If goals are important, as the pin problem data suggest they are, then similarity of goals should have increased the chance that the pin training problem would have been retrieved when it had the same goal as a problem requiring a fork solution. The more analogous a fork problem was to the pin training problem, the slower should have been the participants' responding. Yet there were no differences between response times for different types of fork problems, no matter how similar they were to the pin training problem. Such a finding appears to be consistent with goals being a part of the process, rather than a separate stage. However, this requires more examination.

ACKNOWLEDGMENTS

I wish to thank Missy Schreiner for assistance in this research, and Regina Vollmeyer for comments on an earlier draft of this paper. Final preparation of this paper was assisted by a stipend from the Humboldt Foundation.

REFERENCES

Broadbent, D. E (1958). *Perception and communication.* London: Pergamon Press.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence, 41*, 1-63.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Analogy, similarity, and thought* (pp. 199-241). New York: Cambridge University Press.

Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy. *American Psychologist, 52*, 32-34.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45-56.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13, 295-355.

Holyoak, K. J., & Thagard, P. (1997). The analogy mind. *American Psychologist, 52*, 35-44.

Klein, G. A. (1986). *Analogical decision making* (ARI Research Note 86-102). US Army, Research Institute for the Behavioral and Social Sciences.

Posner, M. I. (1986). *Chronometric explorations of mind.* New York: Oxford University Press.

Schank, R. C. (1982). *Dynamic memory.* Cambridge: Cambridge University Press.

Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology, 31*, 307-346.

Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal learning and Verbal Behavior, 3*, 449-459.

MAPPING CONCEPTUAL AND SPATIAL SCHEMAS

Merideth Gattis
Max Planck Institute for Psychological Research
Leopoldstrasse 24
80802 Munich, Germany
gattis@mpipf-muenchen.mpg.de

ABSTRACT

Three experiments used an artificial sign language to investigate whether the mapping of verbal statements to spatial schemas is constrained by similarity of relational structures. In Experiment 1 adults were shown diagrams of hand gestures paired with locative statements, and asked to judge the meaning of new gestures. In Experiment 2, adults were asked to make similar judgments with active declarative statements. In Experiment 3, the artificial signs were paired with conjunctive and disjunctive relations. Results of all three experiments indicate that adults choose a physical object to represent a conceptual element and a physical relation to represent a conceptual relation. These results corroborate the structure-driven mapping patterns found in previous studies of visual reasoning, and provide further support that visual reasoning is based on general cognitive constraints on mapping concepts to space.

From the early likening of sound to waves to the more recent comparison of armies and rays, many analogies intertwine spatial and conceptual components so tightly that it seems difficult to unravel how they first came together. Perhaps this melding is one reason why many investigations of analogy have involved comparison of problems which may be presented verbally or visually without asking how the two forms of representation are related. The question of how spatial and conceptual information are linked is important not only for understanding analogy, but also for understanding how spatial structures influence the use of diagrams and models in reasoning (Glasgow, Narayanan, & Chandrasekaran, 1995), the structure of languages (Bloom, Peterson, Nadel, and Garrett, 1996), and perhaps even the origins of abstract cognitive abilities (Pinker, 1989).

MAPPING CONCEPTS TO SPACE

Research on reasoning with spatial representations suggests two possible principles governing the mapping of conceptual and spatial schemas (Gattis, 1997). Consistent mappings may derive from meaningful associations, such as the association between "more" and "up," or from structure-driven mapping, matching conceptual and spatial schemas based on structural similarities.

Association-based Mapping

Associations between physical aspects of the world and conceptual aspects of experience are frequently reflected in language, such as the association between "more" and "up" reflected in metaphorical expressions like "My income rose last year" (Lakoff & Johnson, 1980, pp.15-16). Such associations may influence how people map conceptual schemas to spatial schemas. Research on children"s graphic constructions indicates that when asked to place stickers on a piece of paper to represent increases, children representing quantitative increases in a vertical direction are more likely to place the lowest level (i.e. "a small amount") at the bottom of the page and the highest level (i.e. "a really big amount") at the top of the page (Gattis, 1997; Tversky, Kugelmass, & Winter, 1991). Similarly, adults asked to map relational terms to vertical or horizontal lines mapped "above" and "below," "better" and "worse," and "more" and "less" most often to a vertical axis, with the first term of each pair at the top, and the latter term at the bottom (Handel, DeSoto, & London, 1968).

Structure-driven Mapping

Association-based mappings appear to be inadequate, however, for explaining the diverse interactions of mapping patterns in visual reasoning. The direction and strength of some mapping patterns are not easily explained by association-based mapping, such as the tendency to map "steeper" and "faster" reported not only in adults but also in young children with no graphing experience (Gattis, 1997; Gattis & Holyoak, 1996). Our experience in the physical world is as likely to lead to an association between "steeper" and "slower" as between "steeper" and "faster," since steeper hills lead to slower rates of travel uphill and faster rates of travel downhill.

In addition, association-based mappings may come into conflict, and when multiple mappings conflict, some mappings reliably take precedence over others. Gattis and Holyoak (1996) asked adults to reason with graphic constructions which contrasted two natural mappings: the iconic mapping of "up" on a vertical line and "up" in the atmosphere against the metaphoric mapping between steeper slope and faster rate of change. The latter mapping exerted a stronger influence on reasoning performance. Thus a coherent system appears to guide which mapping is used, even if some mappings may be derived from prior associations.

A second explanation for mapping consistencies is that mappings between concepts and space are based on general constraints governing the mapping process, rather than or in addition to specific associations. An example of such a general constraint is the tendency observed in analogical mapping to map two concepts based on structural similarities (Gentner, 1983). Structure-driven mapping is appealing because it can explain reported mapping patterns for reasoning both about quantities and about rates. When Gattis (1997) asked young children to reason about quantity or rate using graph-like diagrams, children"s judgment patterns revealed two highly consistent mappings of concepts to spatial dimensions: quantity was inferred from the height of a line and rate was inferred from the slope of a line. Mapping of concepts to space thus appears to be governed by relational structure. Young children mapped quantity to height —structurally similar because they are both relations between elements — and rate to slope — structurally similar because they are both relations between relations.

Mapping Relational Structure

The studies reported here focus on very simple relational structures — elements and relations between elements — to further explore how relational structure is defined in conceptual and spatial schemas. Three experiments used an artificial sign language to investigate whether adults" conceptual interpretations of completely novel spatial schemas would also be characterized by structure-driven mapping. If visual reasoning is indeed based on mapping relational structures from conceptual to spatial schemas, judgment patterns ought to reflect mapping of conceptual elements to physical objects and conceptual relations to physical relations.

Experiment 1: Relational StructureIn Locative Statements

Experiment 1 examined whether relational structure influences mapping of locative statements to spatial schemas by asking adults to interpret an artificial sign language in a three phase procedure. The first phase assigned a specific meaning to each hand, as seen in Figure 1. The second phase paired two signs made with the right hand with two simple locative statements involving the object represented by the right hand. The two signs were touching the right ear with the right hand and touching the left ear with the right hand, as seen in Figure 2. These two signs were intentionally ambiguous: the assignment of one locative to each sign leaves open whether it is the object touched by the hand (right ear, left ear) or the relation of the hand to body (ipsilaterial, contralateral) that carries meaning. The third phase introduced two complementary signs made with the left hand (touching the left ear with the left hand and touching the right ear with the left hand, illustrated in Figure 3), and asked participants to judge which of two new locative statements was represented by each new sign.

The four locative statements used in Experiment 1 were "Mother is in the car," "Mother is in the office," "Father is in the car," and "Father is in the office." Two types of relational structure were contrasted by varying which aspect of the statement was clearly mapped and which aspect of the statement was ambiguously mapped. This was accomplished by manipulating which aspect of the locative statement was assigned to the hands. The meanings assigned to the right and left hands were either "car" and "office," or "mother" and "father." For those participants for whom "car" and "office" were assigned to the hands, the subjects of the locative statements introduced in the second phase ("mother" and "father") were unassigned and therefore ambiguously mapped. In contrast, for those participants for whom "mother" and "father" were assigned to the hands, the locative predicates[1] ("car" and "office") were unassigned and therefore ambiguously mapped.

The expectation was that structure-driven constraints on mapping conceptual to spatial schemas would lead people to map the unassigned portion of the statement to a structurally similar aspect of the accompanying sign. Assigning "car" and "office" to the hands leads to ambiguously mapped subjects, "mother" and "father," and participants in this condition were expected to map those subjects to physical elements of the sign (the right and left ears). Assigning "mother" and "father" to the hands leads to ambiguously mapped locative predicates, "car" and "office," and participants in this condition were expected to map predicates to a physical relation in the sign (the ispilateral and contralateral relations of the arm to the rest of the body). These two mapping patterns were then predicted to lead to opposite judgment patterns in the final phase.

METHOD

Participants. One hundred and thirty-eight students from the University of Technology, Chemnitz and the University of Munich participated in Experiment 1. Experiment materials were distributed and completed during a psychology class, and participation was voluntary. Approximately half of the students were randomly assigned to each of the two conditions.

Two experimental questions at the end served as a consistency measure, and those participants who did not answer the two questions consistently were not included in the analyses (see results section for details). Three subjects from each of the two conditions did not answer these two questions consistently and were discarded from the analyses, resulting in 72 subjects in the S condition and 60 subjects in the R condition (the S and R conditions are explained below).

Procedure and Design. Participants were given a booklet of three sheets of paper stapled together. Each page contained two illustrations, each accompanied by a simple declarative statement. Each illustration and the accompanying statement occupied approximately half a page, and the materials were organized vertically so that the first illustration occupied the top half of the first page, the second illustration occupied the bottom half of the first page, and so on. The instructions were, "Please read the following carefully. At the end you will be asked questions about it."

On the first page were two drawings, first a drawing of a character extending his right hand, and then the same character extending his left hand (see Figure 1). Above each drawing was a sentence "This hand means ______." For half of the participants, the last blank was filled with the words "mother" and "father," with the order counterbalanced so that for half of those participants the right hand meant, "mother," and for half the right hand meant, "father." For the other half of the participants, the last blank was filled with the words, "car" and "office," with the order of car and office counterbalanced in the same way. Varying the assignment of meaning to the hands was the primary experimental manipulation, because it had the effect of varying which portion of the locative statements introduced in the second phase was unassigned. When the hands meant "car" and "office," the unassigned portion of the locative statements was the subject ("Subject varies" or S condition). When the hands meant "Mother" and "Father," the unassigned portion of the locative statements was the relational predicate ("Predicate varies" or R condition). All other manipulated variables were counterbalancing for assignment of meaning to left and right hands and assignment of meaning to each sign.

---

[1] For both conditions, the phrase "is in the" was introduced in the second phase and therefore was part of the unassigned meaning. The phrase gets parsed with its object ("is in the car" or "is in the office") however, thus constituting a predicate phrase.

On the second page were two new drawings, showing the same character first touching his right ear with his right hand, and then touching his left ear with his right hand, or in the opposite order. Above each drawing was a sentence. For the S condition, the two sentences were "This means "Mother is in the car"" and "This means "Father is in the car,"" or "This means "Mother is in the office"" and "This means "Father is in the office,"" depending on the counterbalancing of assignment of "car" and "office" to right and left hands in the first phase. For the R condition, the two sentences were "This means "Mother is in the car"" and "This means "Mother is in the office,"" or "This means "Father is in the car"" and "This means "Father is in the office,"" depending on the counterbalancing of assignment of "Mother" and "Father" to right and left hands in the first phase. Note that in the S condition, the subject of the sentence varies, and in the R condition, the object of the locative predicate varies, but all participants received the same two signs. For both conditions, the two signs paired with these locative statements are ambiguously mapped: it is not clear whether it is the object touched by the hand (right ear and left ear) or the relation of the hand to body (ipsilaterial and contralateral) that means "Mother" and "Father," or "car" and "office."

On the third page were two new drawings of the same character making the complementary signs with his left hand, first touching his left ear with his left hand, and then touching his right ear with his left hand, or in the opposite order. Above each drawing was the question, "What does this mean?", and below each drawing were two sentences, with the instruction, "Circle the answer that fits best." For the S condition, the two sentences read, ""Mother is in the office" *OR* "Father is in the office,"" or ""Mother is in the car"*OR* "Father is in the car,"" depending on the counterbalancing of assignment of "car" and "office" to right and left hands in the first phase. For the R condition, the two sentences read, ""Father is in the car"*OR* "Father is in the office,"" or ""Mother is in the car"*OR* "Mother is in the office,"" depending on the counterbalancing of assignment of "Mother" and "Father" to right and left hands

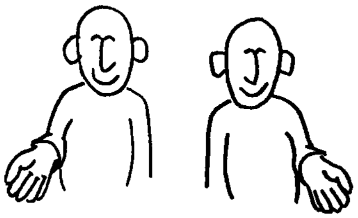


Figure 1. In the first phase, text accompanying these drawings assigned a particular meaning to each hand.

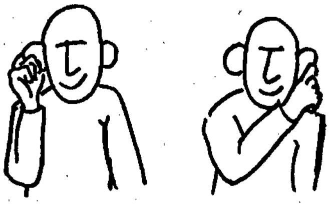


Figure 2. In the second phase, two signs made with the right hand were accompanied by two locative statements (Experiment 1), two active declarative statements (Experiment 2), or two conjunctive or disjunctive statements (Experiment 3).

in the first phase. For both conditions the order of these two sentences was also counterbalanced. As in the second phase, in the S condition, the subject of the sentence varies, and in the R condition, the object of the locative predicate varies, but all participants received the same two illustrations of signs made with the left hand (see Figure 3).

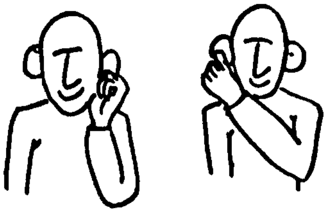


Figure 3. In the third phase, two signs made with the left hand, complementary to those previously shown with the right hand, were used to probe conceptual-spatial mappings.

RESULTS AND DISCUSSION

Asking participants to judge the meaning of two signs with the left hand allowed for a consistency measure, in that the two signs logically ought to have two different meanings. Participants who circled the same locative statement for both signs were therefore considered to have given inconsistent answers, and were discarded from the analyses.

The answers given by the remaining participants were then coded by whether the unassigned meaning was mapped to the ears (object-based or O mappings) or to the ipsilateral and contralateral bodily relations (relational or R mappings). This was easily derived by comparing the circled statements for each sign with the statement-sign pairs on the previous page. For example, a participant in the S condition (subject varies) circled "Father is in the office" as the meaning of the left-hand-to-right-ear sign, and "Mother is in the office" as the meaning of the left-hand-to-left-ear sign. Since it was already known that for this person "car" and "office" were assigned to right and left hand respectively, comparing these judgments with the statement-sign pairings on the second page, where the right-hand-to-left-ear sign was labelled "Mother is in the car" and the right-hand-to-right-ear sign was labelled "Father is in the car," indicated that this person mapped the varying subjects, "mother" and "father" to the left and right ears, respectively. This answer was coded as an object-based (O) mapping. Had this same participant selected "Mother is in the office" and "Father is in the office" for the left-hand-to-right-ear and left-hand-to-left-ear signs respectively, that would have been coded as a relational (R) mapping.

Compare that with the answers given by a participant in the R condition (predicate varies). This person circled "Father is in the car" " as the meaning of the left-hand-to-right-ear sign and "Father is in the office" as the meaning of the left-hand-to-left-ear sign. Since it was already known that for this person "mother" and "father" were assigned to right and left hand respectively, comparing these judgments with the statement-sign pairings on the second page, where the right-hand-to-left-ear sign was labelled "Mother is in the car" and the right-hand-to-right-ear sign was labelled "Mother is in the office," indicated that this person mapped the varying predicates "is in the car" and "is in the office" to the contralateral and ipsilateral bodily relations, respectively. This answer was coded as a relational (R) mapping. Had this same participant selected "Father is in the office" and "Father is in the car" for the left-hand-to-right-ear and left-hand-to-left-ear signs respectively, that would have been coded as an object-based (O) mapping.

	R	O	total
Subject varies (Mother/Father)	26	46	72
Predicate varies (is in the car/is in the office)	43	18	60

Table 1. Frequencies of relational (R) and object-based (O) mappings for conditions in which subject varies and in which predicate varies in Experiment 1.

The frequency of relational and object-based mappings were then compared. As can be seen in Table 1, the assignment of meaning to an ambiguous sign was determined by which aspect of the locative statement was unassigned. Participants in the S condition (subject varies) were more likely to make object-based mappings, whereas participants in the R condition (predicate varies) were more likely to make relational mappings. These two patterns of response for the S condition and the R condition were significantly different $C^2(1, N = 133) = 15.64$, $p < .001$. The overall frequency of the two mapping patterns were approximately the same: combining the two experimental conditions, participants chose relational and object-based mappings with similar frequency.

The results of Experiment 1 were consistent with structure-driven mapping of conceptual to spatial schemas. When the subject of the locative statement was unassigned, participants mapped the unassigned subjects to physical objects — the right and left ears. When the predicate of the locative statement was unassigned, participants mapped the unassigned predicates to physical relations — the ipsilateral and contralateral relation of the arm to the body.

Experiment 2: Relational Structure In Active Declarative Statements

The results of Experiment 1 indicated that relational structure influences the mapping of locative statements to novel spatial schemas. Experiment 2 investigated whether mapping active declarative statements to novel spatial schemas would reveal the same pattern of structure-driven mapping. Experiment 2 used the same diagrams and three phase procedure as Experiment 1. Whereas Experiment 1 paired signs with simple locative statements, such as "Mother is in the office" and "Mother is in the car," Experiment 2 paired signs with active declarative statements.

The active declarative statements used in Experiment 2 were about animal characters performing some action toward another animal character: for example, "Monkey visits Mouse," and "Monkey bites Mouse." As in Experiment 1, different types of relational structure were contrasted by varying which aspect of the statement was clearly mapped and which aspect of the statement was ambiguously mapped by assigning a particular aspect of the statements to the hands. The hands always signified two animals, assigned during the first phase using the same procedure as Experiment 1.

In the statement pairs introduced in the second and third phases, either the subject varied (S condition), the relation varied (R condition), or the object varied (O condition). When the subject varied (S condition), the meanings assigned to the hands always became the objects of the action, and two subjects were introduced. The relation was constant for an individual participant, but varied between-subjects. For instance, a participant in the S condition for whom "Monkey" had been assigned to the right hand in the first phase, in the second phase might have read the statements, "Mouse visits Monkey" and "Bear visits Monkey," each paired with a diagram of a sign made with the right hand (see Figure 2).

When the relation varied (R condition), the meanings assigned to the hands became the subjects of the action, and two relations were introduced. A participant in the R condition for whom "Monkey" had been assigned to the right hand in the first phase, in the second phase might have read the statements, "Monkey visits Mouse" and "Monkey bites Mouse," also paired with the right hand signs (Figure 2).

When the object varied (O condition), the meanings assigned to the hands became the subjects of the action, and two objects were introduced. A participant in the O condition for whom "Monkey" had been assigned to the right hand in the first phase, in the second phase might have read the statements, "Monkey visits Mouse" and Monkey visits Bear," each paired with a right hand sign (Figure 2).

As in Experiment 1, in Experiment 2 the expectation was that structure-driven constraints on mapping conceptual to spatial schemas would lead people to map the unassigned and varying portion of the statement to a structurally similar aspect of the accompanying sign.

Varying subjects and varying objects were expected to lead to more object-based mappings, assigning meaning to the right and left ears. Varying relations, in contrast, were expected to lead to more relational mappings, assigning meaning to the ispilateral and contralateral relations of the arm to the rest of the body. Experiment 2 thus manipulated three different aspects of statements, which when mapped to spatial schemas were expected to lead to two distinct mapping patterns: both varying subjects and varying objects should lead to object-based mapping of conceptual to spatial schemas, whereas varying relations should lead to relational mapping of conceptual to spatial schemas. These two mapping patterns were again predicted to lead to opposite judgment patterns in the final phase.

METHOD

Participants. One hundred and fifty-four students from the the University of Munich participated in Experiment 2 during psychology classes. Participation was voluntary. Approximately one-third of the students were randomly assigned to each of the three conditions.

As in Experiment 1, two experimental questions at the end served as a consistency measure. Six subjects in the S condition, and two subjects in each of the remaining conditions, did not answer these two questions consistently and were discarded from the analyses, resulting in 32 subjects in the S condition and 58 subjects in the R condition, and 52 subjects in the O condition.

Procedure and Design. The procedure and materials were nearly identical to those of Experiment 1, with the change that the statements paired with signs in the second and third phases were active declarative statements, and there were three experimental conditions: subject varying (S condition), relation varying (R condition), and object varying (O condition).

On the first page of the experimental booklet were two drawings of a character extending his right and then left hand (see Figure 1), paired with the sentences "This hand means (Animal1)," and "This hand means (Animal2)." Whereas in Experiment 1 varying the assignment of meaning to the hands was related to the primary experimental manipulation, in Experiment 2 it was independent. Both the subjects and the objects of the active declarative statements were animal characters, and any animal could be assigned to the hands. Four animals were chosen — Monkey, Elephant, Mouse, and Bear — and a random ordering of these four was created. Three new orderings were created by rotating the the list. The other three orders were thus: Elephant, Mouse, Bear, and Monkey; Mouse, Bear, Monkey, and Elephant; and Bear, Monkey, Elephant, and Mouse. The first position in the list was Animal1, the second position in the list was Animal2, and so on. These four orders were counterbalanced between subjects.

On the second page of the booklet were two more drawings of the same character touching his right ear with his right hand, and touching his left ear with his right hand (see Figure 2), with the order of these two drawings counterbalanced across subjects. Above each drawing was a sentence. For the S condition, the two sentences were of the form, "This means "Animal3 R-action Animal1"" and "This means "Animal4 R-action Animal1."" The relation (R-action) was either "visits" or "bites" and was counterbalanced across subjects. For the R condition, the two sentences were of the form, "This means "Animal1 R-action1 Animal3"" and "This means "Animal1 R-action2 Animal3."" As with the S condition, the relations were "visits" and "bites," and the order of these two relations was counterbalanced across subjects. For the O condition, the two sentences were of the form, "This means "Animal1 R-action Animal3"" and "This means "Animal1 R-action Animal4,"" and again the relation was either "visits" or "bites" and was counterbalanced across subjects.

On the third page of the booklet were two drawings of the same character touching his left ear with his left hand, and touching his right ear with his left hand (see Figure 3), with the order of these two drawings counterbalanced. Above each drawing was the question, "What does this mean?", and below each drawing were two sentences, with the instruction, "Circle the answer that fits best." For the S condition, the two sentences were of the form, "This means "Animal3 R-action Animal2" *OR* This means "Animal4 R-action Animal2."" For the R condition, the two sentences were of the form, "This means "Animal2 R-action1 Animal3" *OR* This means "Animal2 R-action2 Animal3."" For the O condition, the two sentences were of the form, "This means "Animal2 R-action Animal3" *OR* This means "Animal2 R-action Animal4."" For each condition, the order of the two sentences was counterbalanced across subjects. Note that the two sentences are identical to those introduced in the second phase, with the single change that Animal2, which has already been assigned to the left hand, is substituted for Animal1, to correspond to the left-handed actions in the drawings.

RESULTS AND DISCUSSION

As in Experiment 1, participants who circled the same locative statement for both signs were considered to have given inconsistent answers and were discarded from the analyses. The answers given by the remaining participants were then coded by whether the unassigned meaning was mapped to the ears (object-based or O mappings) or to the ipsilateral and contralateral bodily relations (relational or R mappings), in the same manner described in Experiment 1, and the frequency of relational and object-based mappings were then compared. The mapping pattern varied significantly between conditions $C^2(2, N = 142) = 16.49$, $p < .001$. As can be seen in Table 2, participants in the S condition (subject varies) and the O condition (object varies) were more likely to make object-based mappings, whereas participants in the R condition (predicate varies) were more likely to make relational-mappings.

	R	O	total
Subject varies (Bear/Elefant/Maus/Monkey)	11	21	32
Relation varies (bites/visits)	36	22	58
Object varies (Bear/Elefant/Maus/Monkey)	13	39	52

Table 2. Frequencies of relational (R) and object-based (O) mappings for conditions in which subject varies, in which action predicate varies, and in which object varies in Experiment 2.

Experiment 3:
Relational Structure
In Conjunctive and Disjunctive Statements

Experiment 3 investigated whether the same type of relational structures are revealed in the mapping of conjunctions and disjunctions to artificial signs, using the same diagrams and three phase procedure as the previous experiments. The statements paired with signs in Experiment 3 were simple conjunctions and disjunctions of animal characters, such as "Monkey and Mouse," and "Monkey or Mouse."

As in Experiments 1 and 2, different types of relational structure were contrasted by varying which aspect of the statement was clearly mapped and which aspect of the statement was ambiguously mapped by assigning a particular aspect of the statements to the hands. As in Experiment 2, the hands always signified two animals, assigned during the first phase of the experiment. In the statements paired with signs in the second and third phases of the experiment, either the first animal (S condition), the second animal (O condition), or the relation between them (R condition) varied.

When the first animal varied (S condition), two new animals were paired with the animal previously assigned to the right hand, by either a conjunctive relation ("and") or a disjunctive relation ("or"). The relation was constant for an individual participant, but varied between-subjects. To compare with the examples described in Experiment 2, a participant in the S condition of Experiment 3 for whom "Monkey" had been assigned to the right hand in the first phase, in the second phase might have read the statements, "Mouse and Monkey" and "Bear and Monkey," each paired with a diagram of a sign made with the right hand (see Figure 2).

The O condition was similar to the S condition, with the difference that the animals assigned to the hands occupied the first position in the statements, and the two new animals occupied the second position in the statements. For example, a participant in the O condition for whom "Monkey" had been assigned to the right hand in the first phase, in the second phase might have read the statements, "Monkey and Mouse" and "Monkey and Bear," each paired with a right hand sign (Figure 2).

When the relation varied (R condition), both conjunctive ("and") and disjunctive ("or") relations were introduced. A participant in the R condition for whom "Monkey" had been assigned to the right hand in the first phase, in the second phase might have read the statements, "Monkey and Mouse" and "Monkey or Mouse," also paired with the right hand signs (Figure 2).

As in the previous two experiments, the expectation was that structure-driven mapping would pair the unassigned and varying portion of the statement with a structurally similar aspect of the accompanying sign. Varying animals, whether in the first position or the second position, were expected to be mapped to the ears, an object-based mapping. Varying relations, in this case "and" and "or," were expected to be mapped to the ispilateral and contralateral relations of the arm to the rest of the body. These two mapping patterns were again predicted to lead to opposite judgment patterns in the final phase.

METHOD

Participants. One hundred and six students from the the University of Munich and the University of Chemnitz participated in Experiment 3 during psychology classes. Participation was voluntary. Approximately one-third of the students were randomly assigned to each of the three conditions.

As in Experiments 1 and 2, two experimental questions at the end served as a consistency measure. Four subjects in the S condition, one subject in the R condition, and three subjects in the O condition did not answer these two questions consistently and were discarded from the analyses, resulting in 44 subjects in the S condition and 32 subjects in the R condition, and 30 subjects in the O condition.

Procedure and Design. The procedure and materials were nearly identical to those of Experiment 2, with the change that the statements paired with signs in the second and third phases were conjuntive pairs, disjunctive pairs, or both. As in Experiment 2, there were three experimental conditions: first animal varying (again called the S condition, to allow easy comparison with Experiment 2, despite the fact that in Experiment 3 the first animal is not the subject of a sentence), relation varying (R condition), and second animal varying (again called the O condition, despite the fact that the second animal is not the object of a sentence).

RESULTS AND DISCUSSION

As in the two previous experiments, participants who circled the same statement for both signs were considered to have given inconsistent answers and were discarded from the analyses. The answers given by the remaining participants were then coded by whether the unassigned meaning was mapped to the ears (object-based or O mappings) or to the ipsilateral and contralateral bodily relations (relational or R mappings), and the frequency of relational and object-based mappings were then compared. The mapping pattern varied significantly between conditions $X^2(2, N = 121) = 10.21$,

	R	O	total
First animal varies (Bear/Elefant/Maus/Monkey)	15	35	50
Relation varies (and/or)	20	11	31
Second animal varies (Bear/Elefant/Maus/Monkey)	14	26	40

Table 3. Frequencies of relational (R) and object-based (O) mappings for conditions in which the first animal varies, in which relation varies, and in which the second animal varies in Experiment 3.

$p < .006$. As can be seen in Table 3, participants in the S condition and the O condition were more likely to make object-based mappings, whereas participants in the R condition were more likely to make relational-mappings.

General Discussion

The three experiments reported here used an artificial sign language to investigate whether the mapping of simple statements to spatial schemas is constrained by similarity of relational structures. In Experiment 1 adults were shown diagrams of hand gestures paired with locative statements, and asked to judge the meaning of new gestures. In Experiment 2 and 3, adults were asked to make similar judgments with active declarative statements and conjunctive and disjunctive statements, respectively. Results of all three experiments indicate that adults choose physical objects to represent conceptual elements and physical relations to represent conceptual relations. These results corroborate the structure-driven mapping patterns found in previous studies of visual reasoning, in which relations between elements were mapped together, and relations between relations were mapped together (Gattis, 1997).

The results reported here are also compatible with previous research with signed languages indicating that that in signing space, objects or actors are assigned to a spatial locus (Emmorey, 1996). Interestingly, however, these results also indicate that nouns are not always assigned to spatial loci, but rather the structural role played by a noun determines whether it is mapped to a spatial locus or a spatial relation. In Experiment 1, the nouns "car" and "office" were mapped to the ispilateral and contralateral bodily relations, not to the right and left ears, because they were essential parts of locative relational expressions, "in the car" and "in the office."

One alternative explanation to the structure-driven mapping interpretation suggested here is that adults are mapping roles and movement rather than structure per se when mapping statements to signs. For instance, the ipsilateral and contralateral gestures could be interpreted as movements rather than bodily relations. The tendency to pair the locative expressions "in the car" and "in the office" with the ipsilateral and contralateral relations could be seen to emphasize the movement toward a location, rather than a bodily relation. This explanation is a variant of the association-based mapping hypothesis, and assumes that people associate movement of the arms with movement to a location, or the movement of an action such as "bite" or "visit." If people perceive and map the movement path, however, we would also expect that the movement marks the grammatical roles of subject and object, as in signed languages (Emmorey, 1996). When movement is used to represent an action in ASL, such as "The dog bites the cat," the direction of the movement marks the grammatical roles of subejct and object: the subject is the starting location, and the object is the end location. Were adults simply mapping locations and actions to the signs shown here by mapping location and action to a movement path, we would expect to find stronger mapping patterns for those situations in which the object of the statement was mapped to the ears compared to those in which the subject of the statement was mapped to the ears. In Experiment 2, however, subjects and objects of active declarative statements were mapped with equal freqency to the the ears, indicating that perceived movement did not play an important role in adults" mapping of conceptual schemas to spatial schemas.

By introducing a new paradigm for studying the mapping of conceptual and spatial schemas, these experiments also provide an interesting task for studying relational structure in language. The results of all three experiments indicate that adults asked to interpret this artificial sign language choose a distinct mapping pattern, either object-based or relational, to map linguistic structures to hand gestures. Further research might use this paradigm to address the relational structure underlying linguistic utterances as well as reasoning schemas.

ACKNOWLEDGEMENTS

This work was funded by the Max Planck Society. I would like to thank Christophe Lardschneider for making the drawings used in these experiments, Marc Buehner for conducting experiments at the University of Technology, Chemnitz, and Denise Parks for assistance in creating materials and conducting experiments at the University of Munich.

REFERENCES

Bloom, P., Peterson, M., Nadel, L., & Garrett, M. (1996). *Language and space*. Cambridge, MA: MIT Press.

Emmorey, K. (1996). The confluence of space and language in signed languages. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp.171-209). Cambridge, MA: MIT Press.

Gattis, M. (1997). Structure-driven mapping in visual reasoning. Manuscript under review.

Gattis, M., & Holyoak, K.J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 231-239.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Glasgow, J., Narayanan, N.H., & Chandrasekaran, B. (1995). *Diagrammatic reasoning*. Menlo Park, CA: AAAI Press.

Handel, S., DeSoto, C.B., & London, M. (1968). Reasoning and spatial representations. *Journal of Verbal Learning and Verbal Behavior, 7*, 351-357.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23*, 515-557.

ANALOGICAL DISTANCE AND PURPOSE IN CREATIVE THOUGHT: MENTAL LEAPS VERSUS MENTAL HOPS

Thomas B. Ward
Department of Psychology
Texas A&M University
College Station, Texas 77843, USA
tbw@psyc.tamu.edu

When people apply existing knowledge to new tasks, the circumstances surrounding that application can vary enormously from one situation to the next. Potentially important variations include the purposes to which the old information is put, the conceptual distance between the old source and the new target domain, and the person's state of knowledge regarding the target. Considering some of these variations can help to provide a broader context for the research I will present and for thinking about knowledge transfer more generally.

Knowledge from a familiar source can be used for the purpose of reasoning about, explaining, or otherwise coming to understand a less familiar target domain, or it can be used to supply the starting point or structuring information needed for the design of novel products, inventions or other tangible artifacts. As a short-hand, these different uses of existing knowledge can be referred to as **explanatory** and **inventive**, respectively.

In terms of conceptual distance, the source and target can come from the same conceptual domain, from related, though nonidentical domains, or from wildly discrepant domains, (e.g., Dunbar, 1997; Vosniadou & Ortony, 1989). For ease of reference, those continuous variations can be labeled loosely with the dichotomous terms **near** and **distant**.

Finally, individuals seeking to apply source knowledge to a target situation may know a great deal or next to nothing about the target. As discussed below, initial knowledge about the structure of the target should be richer in the explanatory than in the inventive case.

A PARTITIONING OF CASES

The explanatory/inventive and near/distant distinctions can be used to partition knowledge transfer situations into several types. For example, classic instances of real-world analogies, particularly those involved in scientific discovery, are typically characterized by the use of a well-known, but conceptually distant source domain to explain or understand a relatively less familiar target domain. An oft noted instance of this type of **distant/explanatory** analogy is Rutherford's comparison between the familiar structure of a solar system and the (then) relatively unknown structure of the atom. Another less noted, but equally striking instance is Kepler's analogy between the properties of light and a hypothetical motive power of the sun which he invoked to try to explain planetary motion (Gentner, Brem, Ferguson, Wolff, Markman, & Forbus, 1997).

Distant sources are also reported to serve the purpose of envisioning, designing, and producing novel inventions. A frequently cited instance of this type of **distant/inventive** analogy is the role of burrs in the invention of velcro. According to the story, when velcro's inventor, George de Mestral, used a microscope to examine burrs that had attached to his clothing, he noticed that they were collections of miniature "hooks" that had locked into the "eyes" in the cloth of his pants and socks. Mestral used that knowledge to design a similar system of miniature hooks-and-eyes that could be used as a fastener.

Recent observations of the activities of molecular biology laboratory groups have also identified a preponderance of **near/explanatory** analogies, which involve the use information from either the same domain in a different context, or a closely related source domain to understand the target domain (e.g., Dunbar, 1997). Instances of these types of analogies identified by Dunbar include a mapping from how HIV operates in an in vivo context to how it works in an in vitro context, and a mapping between the Ebola and Herpes viruses.

To complete the set, the world is replete with instances of **near/inventive** analogies in which individuals stay within a domain, but push its boundaries by envisioning and bringing to fruition novel exemplars of that domain.. The term "inventive" here is not used to restrict these types of analogies to the acts associated with producing patentable inventions, but rather to contrast them with those analogies designed primarily to explain or understand a phenomenon. Thus, when an engineer designs a new gear, a novelist crafts a new unlikely hero, or a country singer pens a new ballad, their creative activities can all be seen as instances of near/inventive analogy use. Examples of this type of activity abound, and they include specific cases of invention, such as Thomas Edison's patterning of his electric light distribution system after the existing gas light distribution system of his day (Friedel & Israel, 1979), and Eli Whitney's use of the existing charka as the basis for his cotton gin (Basala, 1978). They also include more generic tendencies, such as science fiction writers' reliance on Earth animals as the bases for their imaginary extraterrestrials (Ward, 1994), and architects' reliance on specific instances of prior buildings to accomplish particular goals in the design of new buildings (see e.g., Kolodner, 1997).

MENTAL LEAPS, MENTAL HOPS, MAPPING AND ACCESS

Considerable research has focused on the use of analogy in reasoning and explanation, and, at least from the examples that have been described most often, much attention has been given to distant analogies. In contrast, the current presentation will focus primarily on the sorts of products that emanate from near/inventive uses of existing knowledge, with a particular emphasis on the retrieval of highly representative domain exemplars as sources of information. However, it will also briefly attempt draw out connections to more distant and explanatory types of transfer, and to delineate some of the potential variations in goals and outcomes across the situations. To what extent is the transfer of old knowledge to new situations governed by similar principles across the range of conceptual distances and purposes?

As one possible difference across situations, it is reasonable to postulate that distant analogies are more likely to be associated with extraordinary forms of creativity, whereas near analogies are more likely to be associated with everyday, relatively small creative increments. If distant analogies are seen as creative "mental leaps" (e.g., Holyoak & Thagard, 1995), intra-domain conceptual extensions might be better seen as creative "mental hops," with less deviation from the source and more attributes preserved. That is, because the objects from distant domains will differ greatly in their superficial properties while at the same time participating in comparable relations, only the latter will tend to be mapped (Gentner, 1989) across distant domains.

In contrast, because instances from the same or close conceptual domains will share superficial as well as deeper similarities, those surface properties are more likely to be preserved in the near than in the distant case. Put differently, the new concept that results from the analogy process will generally diverge less from the old ones in near than in far analogies. Near analogies reflect more of a literal similarity between the source and target (e.g., Gentner, 1989), they may represent smaller conceptual changes between the old and new ideas, and thus may be seen as less dramatically creative.

Having linked near analogies to smaller creative advances, however, I hasten to add that this in no way diminishes their importance. Human progress is certainly much indebted to the basic propensity to innovative in small incremental steps that diverge only slightly from what has come before (see e.g., Basala, 1978).

It is important, too, to distinguish the conceptual distance between old and new ideas from the broader impact of those new ideas. For instance, Edison's lightbulb differed only slightly in basic form from several less successful patented versions that preceded it. Yet, the end result of widely available electric light had a dramatic effect on society. Thus, it represented a small hop from what had come before conceptually, but a giant leap in terms of its impact on the world.

Another difference across the types of situations is that the inventive case seems to imply less initial knowledge about the structure of the target, and consequently, a more limited role for an initial mapping between the source and target. Unlike the case of explanatory analogies that presumably arise because there are observations and some amount of knowledge about a target domain that call for further explanation, the "targets" or products of inventive analogies often do not exist until they are created via the projection of structure from the source.

For example, observations about planetary motion existed before Kepler applied knowledge about light to explain or understand those phenomena, whereas the concept of velcro did not exist, even in rudimentary form prior to de Mestral realizing that the structure of burrs could be adapted to produce a reusable fastener. Results from experiments on specific disease processes existed to be explained by near analogies to other known disease processes (Dunbar, 1997), whereas the cotton gin, as a specific product, did not exist until Whitney applied knowledge from its immediate predecessor, the charka, to develop it.

Because the target, **perse**, tends to come into being in the inventive case as a result of the analogical process, determining the mapping between source and target domains is somewhat simplified relative to the explanatory case in which the relational structures of the source and target must be structurally aligned to produce an effective analogy (e.g., Gentner & Markman, 1997). This is not to say that the goals or desirable properties of inventions, story lines, villains, buildings, and so on are not specified in advance or that they play no role in adapting the structure of the source knowledge, but simply that mapping between domains is minimized and projection is emphasized. Inventive analogies seem to reflect, not so much a process of comparison of structures as they do a process of projecting or instantiating a known structure in a novel way.

Although mapping may be minimized, a crucial issue for inventive analogies (as well as explanatory ones) is to characterized how people **access** the source information. What factors determine the retrieval of the information that will serve as the basis for the structure of the novel product? Here too, there may be differences across situations.

Similarity of surface level and structural properties between the target and source is widely acknowledged as being crucial to retrieving sources in explanatory analogical reasoning (see e.g., Dunbar, 1997; Gentner, 1989; Holyoak & Thagard, 1989; 1997; Ross, 1989). However, in the inventive case, the target only exists after the fact, and similarity to the source may be better seen as the **consequence** rather than the **cause** of retrieving a particular source. Alternatively, if the goals for the novel product are well-enough specified, and the person's knowledge is indexed in a way to allows access to previous cases that have satisfied those goals, goal-relatedness might drive retrieval in the inventive case (see, e.g., Kolodner, 1997).

Beyond similarity to the target and the capacity to satisfy the goals for the target, retrieval of source information may well be determined primarily by the properties of the source domain itself as well as more general conceptual processing tendencies, such as a reliance on the basic level of categorization. Without a rich target representation driving the retrieval of a highly similar source, properties of the source domain itself may take on special importance in determining what gets retrieved and used in the inventive case. In the next sections I describe a series of experiments concerned with the near/inventive use of existing knowledge, and I discuss one particular model that highlights the role of the graded structure of source domains and the retrieval of highly representative instances from those domains.

NEAR/INVENTIVE ANALOGICAL PROJECTION

Because the products of near/inventive creative endeavors are direct outgrowths of the concepts that have come before, they can be expected to share important properties with previous exemplars of those concepts. This is true of real-world accomplishments, such as inventions, art, music, writing, and science (e.g., Basala, 1988; Friedel & Israel, 1986; Weisberg, 1986), as well as laboratory-based performance observed in a variety of generative tasks (e.g., Ward, 1994; Ward & Sifonis, 1997).

As an illustration of a laboratory-based study concerned with the role of existing knowledge in near/inventive, creative generation, Ward (1994) asked college students to imagine, draw, and describe animals that might live on other planets. Despite the fact that the planets were described as being completely different from Earth, Ward found that the students' creations tended to be strongly **analogous** to Earth animals in many respects. At the level of superficial similarity of component elements, they were very likely to possess standard sensory organs, such as eyes, and standard appendages, such as legs that were highly similar in appearance to their counterparts in Earth animals.

At a somewhat deeper level, it is also obvious from the participants' drawings and descriptions that the form of these imagined animals was influenced by the kinds of relational structures that connect the separate elements of Earth animals. That is, the senses and appendages were not simply scattered about randomly, but rather were organized into symmetric wholes within bounded solid forms. Likewise, the component elements of the creations showed a kind of one-to-one correspondence with those of Earth animals in that the individual sense organs and appendages tended to correspond to single matching organs and appendages of Earth animals. Eyes matched eyes and tended to serve only the single function of extracting visual information. Legs matched legs, and tended to serve mobility only.

In addition, although participants did not often state it explicitly, their creations also showed a kind of systematicity. That is, clusters of symmetrically placed elements seemed to play complementary roles within broader goal systems. For example, the eyes serve to collect information about prey, the legs allow an approach to the prey, and the claws provide the capacity to grasp it.

It is important to note, however, that despite their obvious similarity to Earth animals, the imagined animals were only rarely direct replicas of any one specific Earth animal. Thus, they possessed some degree of novelty, while still preserving much of the structure of the source domain of Earth animals.

Although with hindsight, these results are not terribly surprising, it is important to note that living things on other planets could conceivably take any of an infinite variety of forms. There is no reason, in principle, why they would have to resemble Earth animals in their surface form. Nevertheless, people projected many of the characteristic properties of Earth animals onto their imagined extraterrestrials. Similar results have been found with other conceptual domains, such as faces (Bredart, Ward, & Marczewski, in press), and with other age groups, such as young children (Cacciari, Levorato, & Cicogna, 1997).

Taking the properties of the novel creations collectively, they seem to reflect an instance of analogical projection from a well-known source domain (Earth animals), to a relatively unknown target domain (extraterrestrials from planets different from Earth). That is, they were structured by component elements that were projected in way that preserved **structural consistency** or **isomorphism**, as well as a high level of **sytematicity**, which have been identified as important ingredients of analogies (e.g., Gentner, 1989; Gentner & Markman, 1997; Holyoak & Thagard, 1989; 1997).

THE PATH-OF-LEAST-RESISTANCE

To account for the structuring of new ideas by old information, Ward and his collaborators have proposed the path-of-least-resistance model (Ward, 1994; 1995; Ward et al., 1997). According to this model, when people approach the task of developing a new idea, their thinking carries them down paths-of-least-resistance in their conceptual representation of the most relevant knowledge domains. They are assumed to gravitate toward fairly specific (basic level) exemplars of the concept, and to project the properties of those instances onto the novel ideas they are developing. For example, in developing imaginary extraterrestrial animals, rather than remaining at the broad level of "animal" people tend to gravitate toward more specific categories within that domain, and to highly **representative** instances, such as dogs rather than less representative ones, such as iguanas.

Although there are many different measures of representativeness (Barsalou, 1985), the one Ward et al. hypothesized to be most predictive was **Output Dominance**, a measure of how readily instances come to mind. The idea is that the category exemplars that come to mind most readily are the ones most likely to be used as starting points in formulating novel ideas. The rationale is that generating new ideas is cognitively demanding, and people tend to simplify the task by pursuing ideas that come readily to mind.

Ward et al. (1997) have recently provided support for the path-of-least-resistance model. They first determined which exemplars were most representative of the domains of animals, tools, and fruit by having college students list the first 20 items that came to mind for each of those categories. The students' responses were then tabulated to derive Output Dominance scores for each exemplar, that is, the number of students listing each exemplar.

The prediction from the path-of-least-resistance model was that the items that were found to be highest in Output Dominance would be the ones most likely to be used as the basis for novel ideas in tasks of imagination. To test the prediction, Ward et al. (1997) then had different groups of college students imagine animals, tools, and fruit that might exist on other planets. In addition to drawing and describing their creations, the students listed all of the factors they could think of that influenced them during the creation process. Those statements were then examined for references to specific exemplars from those domains (e.g., dogs, hammers, apples, and so on), and across the domains, roughly two-thirds of the participants mentioned relying on such specific exemplars.

References to each exemplar were then tabulated to derive a measure termed **Imagination Frequency**, which is an indicator of the likelihood of any given exemplar being used as a starting point for a novel creation. For instance, of the college students who developed imaginary animals, seven mentioned that they based their creations on dogs, which resulted in dog receiving an Imagination Frequency score of 7. Across all three domains and several procedural manipulations, Imagination Frequency scores were found to be significantly positively correlated (in the .60 range) with Output Dominance scores. That is, the students tended to rely most heavily on those category exemplars that come to mind most readily.

THE UNCONSTRAINED CASE

Although, the global findings reveal that many people retrieve and use specific category instances, and that those instances tend to be highly representative ones, considering variations in the task conditions used by Ward et al. (1997) can provide additional insight into the factors that do and do not affect what people retrieve from the source domains. In the first experiment, participants imagined animals that might live on other planets, but they were given little information about the planets, other than the fact that they were very different from Earth. Participants were free to imagine any creature they could, with no constraints on what it could look like, in what type of environment it might need to survive, and so on. Consequently, it is possible that they gravitated toward specific, highly representative Earth animals in this unconstrained case largely because those animals provided an easy solution to the task at hand; they were quickly retrieved from memory, and they did not violate any specified constraints. But what happens to retrieval when various constraints are imposed or when additional information about the target is given?

DESIGN CONSTRAINTS

In the second experiment of Ward et al. (1997), participants imagined novel tools that might be used by a species of intelligent extraterrestrials. Some participants were given no design constraints, whereas others were asked to imagine tools that could meet the needs of an alien species very unlike humans in that they had no appendages. The idea was that, because manipulation by way of hands is a central property of standard tools, constraining participants to consider such a creature might encourage them to move away from Earth tool exemplars. Alternatively however, the tendency to rely on highly retrievable exemplars of the domain may be strong enough that it remains even when those exemplars would need to be heavily modified to meet task constraints. By this latter view, participants facing the constraint may be just as likely as unconstrained participants to rely on Earth tool models, and they will simply modify those exemplars to meet the needs of the species.

The latter view clearly won out in this particular experiment. Those participants who were constrained to design tools for creatures that had no appendages were just as likely as those who faced no design constraints to retrieve specific instances of Earth tools as starting points, and those retrieved tools were no less likely to be predominantly high in Output dominance. Thus, the relative accessibility of category exemplars can play a powerful role even when other situational constraints are operative. The path-of-least-resistance appears to be a seductive and slippery one.

RETRIEVAL CUES FROM THE TARGET

It is important to note, however, that the representativeness of instances within a domain is flexible rather than rigid (e.g., Barsalou, 1987) Consequently it ought to be possible to bias people to retrieve and make use of particular types of instances. Ward (1994) explored this possibility by providing participants with additional information about the properties of the target. Specifically, different groups of participants were told that the creature to be imagined had feathers, scales, or fur, or they were given no information about its attributes.

The subjects in the "feather" condition were significantly more likely to include wings and beaks as additional features, whereas those in the "scales" condition were significantly more likely to include fins and gills, relative to those in the "fur" or control conditions. More importantly for present purposes, self-reports indicated that participants tended to base their creations on particular instances of known birds, fish, or mammals, in the feather, scales, and fur conditions, respectively. Thus, the different cues provided about the target led to the retrieval of different instances from the source domain of Earth animals, whose properties were then mapped onto the novel entities.

In a subsequent experiment, Ward (1994) examined the interactive effects of two types of information about the target domain on the retrieval and use of specific instances: one was general information about the environment on the creature's planet, and the other was specific attributes of the imagined creature itself. Some participants were told that the planet was composed mostly of molten rock with only a few islands of solid land. To obtain enough food, creatures on the planet needed to be able to travel from one island to the next. Consequently, being able to fly over the molten rock would be an adaptive trait and participants creations were expected to be highly likely to fly.

Other participants were told that the planet had violent winds blowing all around it, from just a couple feet above the surface all the way up to the upper reaches of the atmosphere. Flight on such a planet might be expected to be maladaptive and few flying creatures were expected.

In each planet condition, some participants were given a specific detail about the target creature, namely that it had feathers. Others were told that it had fur.

The most important findings were that a) participants in the Molten-Feather and Molten-Fur conditions were highly likely to design flying extraterrestrials, thus showing a sensitivity to the design constraints in the task, but that b) they appeared to have arrived at those creations by different paths. Participants in the former group were more likely than those in the latter group to produce creatures that were classified as birdlike, and to report basing their creations on specific instances of Earth animals. A plausible account of the findings is that the presence of the cue "feathers" led participants to retrieve exemplars of birds which would have been compatible with the environmental constraints of the Molten planet (i.e., safe travel over the molten areas from one island to the next). In contrast, the cue of "fur" may have led participants to initially retrieve mammalian exemplars which, with the exception of bats, would not possess the desired attribute of flight. Consequently, those exemplars would have been rejected in favor of a different starting point. However, because the cue of "fur" also would have reduced the likelihood retrieving birds, birdlike exemplars would have been unlikely to serve as that next starting point. Such conflicts between retrieved exemplars and desired properties of the target may ultimately have led participants to construct flying creatures on the basis of more general information about flight rather than on the basis of specific known exemplars. Thus, the end-product would be less likely to resemble a bird.

Participants in the Windy conditions were less like to produce flying creatures and less likely than those in the Molten-Feather condition to report a reliance on specific Earth animals. Presumably, those in the Windy-Feather condition might also initially have retrieved birdlike exemplars, but would have rejected or drastically modified them because of their incompatibility with the environmental conditions on the Windy planet.

In general then, the findings suggest that information about the known properties of targets (e.g., feathers) and about other task constraints can interact to determine the probability that people will make use of particular instances from the source domain. Target cues can increase the likelihood of retrieving source instances that have properties that match the cue. When other salient properties of those retrieved exemplars are compatible with the task constraints, people tend to rely heavily on those specific exemplars. When those other properties conflict with task constraints, reliance on specific exemplars can be reduced.

CONSTRAINTS FROM PERCEIVED TASK DEMANDS

It may seem odd that people would gravitate toward highly representative instances when they are trying to be creative. Why not shift to more exotic exemplars, or try to avoid them entirely? One reason that people may not do so in these laboratory tasks is that they perceive the demands of the tasks differently from what we intended. Perhaps they think that they are supposed to use representative exemplars or that highly original products would not be valued.

To examine the role of expectations, Ward et al. (1997) had participants design imaginary fruit under different instructional conditions. Some were told to be creative and others were given no special instructions. The results were straightforward; participants who were given the creativity instructions were just as likely as control participants to rely on highly representative instances of Earth fruit in designing their own creations. Thus, the heavy use of highly representative instances is not due exclusively to perceived demand characteristics. More generally, although expectations will surely matter in some real-world and laboratory situations, category structures may often be powerful enough to produce large effects in spite of those expectations.

ACCESS TO SPECIFIC INSTANCES AND LIMITATIONS ON CREATIVE FUNCTIONING

A particularly intriguing finding is that those participants who report that they base their creations on specific exemplars from the source categories design imaginary products that are rated as showing less originality than those produced by participants who report other types of approaches (Ward, 1994; Ward et al., 1997). That is, their creations diverge less from the characteristic properties of known instances from the source domains. Having brought specific instances to mind, the participants tended to project the properties of those retrieved instances onto their novel creations, with the consequence that those creations showed less innovation than ones produced by participants who adopted different approaches to the task. Thus, it appears that one of the major constraints on generative or creative functioning lies in our natural tendency to rely on previous examples when thinking of novel concepts or ideas. More original products can be expected to result when people avoid the tendency to apply the first available representation to a problem (Ward & Sifonis, 1997).

STRATEGIES AND POPULATION EFFECTS

Relying on specific, highly representative exemplars of a known concept and projecting properties from those exemplars onto novel creations should be seen as strategic choices. More creative individuals may be expected to be more flexible in the use of their conceptual knowledge, better able to avoid reliance on representative instances, and less likely to project characteristic properties from specific exemplars. To examine this possibility we have recently observed the performance of gifted adolescents (who can be hypothesized to possess that cluster of conceptual abilities) in the imaginary fruit task (Ward, Saunders, & Dodds, in press).

The gifted participants showed a balance between flexibility and rigidity in the way they approached the design task. That is, they were less likely than our typical college student samples to rely on specific types of Earth fruit. However, when they did so, they were just as likely to gravitate to the items that come to mind most readily, that is, that are highest in output dominance. The correlations between Imagination Frequency and Output Dominance scores for Earth fruit were nearly identical to those found for college students.

ABSTRACTION AND CONCEPTUAL DISTANCE

The path-of-least-resistance model implies that people should be able to develop more creative ideas by moving back up the path in the conceptual hierarchy to more abstract levels. Properties from any level will be projected onto the novel entity being constructed, but they will be less specific, and thus less constraining at more abstract levels. For example, patterning of a novel creature after a dog might lead to the projection of two eyes placed symmetrically in the head, whereas projection from "living thing" might lead to the projection of "taking in information about the environment," a less constraining property that could be instantiated in an indefinite number of ways.

Moving back up the path might be thought of as enhancing originality by shifting the case from a near analogy to a far one. At a specific level, such as "dog," if the person imports information from yet another source to bolster the originality of the creation, it is likely to be a source in the same superordinate, such as a "cat." The higher the level, the broader the superordinate is and the more distant that other source can be. At a very broad level, such as "living thing," the immediate superordinate might be as broad as "physical entity" which could open the possibility of importing information from a quite distant domain, such as "nonliving thing" (e.g., wheels for appendages). In so doing, the length of the mental hop can be increased so that it more approximates a mental leap.

REFERENCES

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation. **Journal of Experimental Psychology: Learning, Memory, and Cognition, 11**, 629-654.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), **Concepts and conceptual development: Ecological and intellectual factors in categorization** (pp. 101-140). Cambridge: Cambridge University Press.

Basala, G. (1988). **The evolution of technology**. London: Cambridge University Press.

Bredart, S., Ward, T. B., & Marczewski, P. (in press). Structured imagination of novel creatures' faces. **American Journal of Psychology**.

Cacciari, C., Levorato, M. C., & Cicogna, P. (1997). Imagination at work: Conceptual and linguistic creativity in children. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), **Creative thought: An investigation of conceptual structures and processes** (pp. 145-177). Washington, DC: American Psychological Association.

Dunbar, K. (1997). How scientists think: Online creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), **Creative thought: An investigation of conceptual structures and processes** (pp. 461-494). Washington, DC: American Psychological Association.

Friedel, R., & Israel, P. (1986). **Edison's electric light: Biography of an invention**, New Brunswick, NJ: Rutgers University Press.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), **Similarity and analogical reasoning**. Cambridge: Cambridge University Press.

Gentner, D., Brem, S., Ferguson, R., Wolff, P., Markman, A. B., & Forbus, K. (1997). In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), **Creative thought: An investigation of conceptual structures and processes** (pp. 403-460). Washington, DC: American Psychological Association.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. **American Psychologist**, **52**, 45-56.

Holyoak, K. J., & Thagard, P. R. (1989). A computational model of analogical problem solving. In S. Vosniadou & A. Ortony (Eds.), **Similarity and analogical reasoning**. New York: Cambridge University Press.

Holyoak, K. J., & Thagard, P. R. (1995). **Mental leaps**. Cambridge, MA: MIT Press.

Holyoak, K. J., & Thagard, P. R. (1997). The analogical mind. **American Psychologist, 52**, 35-44.

Kolodner, J. L. (1997). Educational implications of analogy: A view from case-based reasoning. **American Psychologist, 52**, 57-66.

Ross, B. H. (1989). Remindings in learning and instruction. In S. Vosniadou & A. Ortony (Eds.), **Similarity and analogical reasoning**. New York: Cambridge University Press

Vosniadou, S., & Ortony, A. (1989). Similarity and analogical reasoning: A synthesis. In S. Vosniadou & A. Ortony (Eds.), **Similarity and analogical reasoning**. New York: Cambridge University Press.

Ward, T. B. (1994). Structured imagination: The role of conceptual structure in exemplar generation. **Cognitive Psychology, 27**, 1-40.

Ward, T. B. (1995). What's old about new ideas? In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), **The creative cognition approach** (157-178). Cambridge, MA: MIT Press.

Ward, T. B., Saunders, K. N. (in press). Creative cognition in gifted adolescents. **Roeper Review**.

Ward, T. B., & Sifonis, C. M. (1977). Task demands and generative thinking: What changes and what remains the same? **Journal of Creative Behavior**, **31**, 245-259.

Ward, T.B., Wilkenfeld, M.J., Sifonis, C.M., Dodds, R.A., & Saunders, K.N. (1997). The role of graded category structure in imaginative thought. Unpublished manuscript.

Weisberg, R. W. (1986). **Creativity, genius and other myths**. New York: Freeman.

REASONING BY ANALOGY AND MEMORY FOR CASES IN THE GAME OF CHESS

Evelyne Cauzinille-Marmèche[1] **and André Didierjean**[2]
Université de Provence and Université Paris V
1 Centre de Recherche en Psychologie Cognitive, UMR CNRS 6561, Université de Provence, 29, avenue Robert Schuman, 13621 Aix-en-Provence. E-mail: evelyne@newsup.univ-mrs.fr
2 Laboratoire Cognition et Communication, Paris V - CNRS, 46 rue Saint-Jacques, 75005, Paris.
E-Mail: Andre.Didierjean@Paris5.sorbonne.fr

Many studies have shown that problem solving by analogy is facilitated when a schema that is potentially applicable to a class of problems is constructed, i.e., when the subject builds an abstract representation structure that includes the goals and subgoals to be reached, the requirements to be met, and the strategy to implement (e.g., Gick & Holyoak, 1983; Cummins, 1992). Nevertheless, a hypothesis recently set forth by many authors (e.g., Brooks, Norman, & Allen, 1991; Gobet & Simon, 1996a, 1996b; Pierce et al., 1996; Anderson, Fincham, & Douglass, 1997) is that several representation structures with different levels of abstraction may in fact coexist, including special cases elaborated at a low level of abstraction. Depending on the extent to which the to-be-solved target problem resembles the corresponding source problems, one or the other of these forms of representation will take precedence. When the target problem is recognized as familiar, an already processed case would be searched for and adapted to it. But when the problem cannot be connected to a known case, an abstract schema would be applied and instantiated (provided, of course, that such a schema exists in long-term memory). There is still little experimental data in support of this hypothesis, but it appears plausible and tempting from the standpoint of cognitive efficiency: it is less costly and faster to adapt a known case, if possible, than it is to systematically reconstruct or recalculate the solving process by applying and instantiating an abstract schema. Moreover, this second hypothesis helps account for the fact that novices (who do not yet have an abstract schema) manage to solve problems when they are very similar to the source (e.g., Reed & Bolstad, 1991).

The experiment reported here provides additional arguments in favor of this hypothesis. Starting from the same source problem, we attempted to lead subjects to construct knowledge at different levels of abstractness. By means of various measures, we then tried to evaluate the specific and/or general knowledge they constructed and used in solving structurally isomorphic problems.

Here is an overall view of the experiment.

Subjects had to find the solution to a particular chess problem: attaining "smothered mate with sacrifice" near the end of a chess game.

The subjects' first task was to understand this source problem. One group of subjects was given an explanation of the problem that focused on the sequence of elementary solving steps. For the second group, the explanation consisted of describing the general principle behind smothered mate with sacrifice and illustrating it with this same source problem. This second experimental condition, likely to trigger self-explanations aimed at linking the example to the general principle, was expected to promote the construction of an abstract schema (e.g., Brown & Kane, 1988).

Next the subjects had to solve two new problems, one that was " like " the source problem both in its structural and visual features, and one that looked different on the surface but was in fact structurally isomorphic. The hypothesis was that subjects given the general solving principle would solve the "unlike" problem better than subjects in the other group. It was also hypothesized that these subjects would do better on the "like" problem, because the terms introduced to explain the solving principle ("smothering", "sacrifice", etc.) and to describe the final goal and the various subgoals were expected to promote the encoding of the specific features of the source problem and thereby facilitate its retrieval and adaptation to the processing of problems recognized as similar (e.g., Catrambone, 1995, 1996).

After solving the two problems (like and unlike), subjects had to recall the source example as accurately as possible. This phase allowed us to determine what specific aspects of the problem were stored in long-term memory. Our hypothesis was that subjects who had been told the general principles underlying the solution would remember the source problem better, since they have payed more attention to the relevant pieces of the chessboard.

Finally, subjects had to order a set of new problems according to how much they resembled the source problem (in terms of the similarity of the solving process). The problems to rank differed from the source problem in their surface features and/or in their structure. The hypothesis was that subjects who had constructed an abstract schema would primarily use structure as a criterion for judging problem resemblance (e.g., Chi, Feltovitch & Glaser, 1981).

This experimental setup — in which a lot of measures allowed us to assess the specificity of the knowledge constructed while others served to evaluate its generality — should provide insight into the representation levels elaborated during the acquisition of micro-expertise, and their use in problem solving.

METHOD

Subjects

Forty-four psychology students (mean age: 23 years 4 months, standard deviation: 11 months) participated in the experiment. All subjects judged themselves to be novices in chess (having played less than once a year) but were familiar with the rules.

Procedure

The experiment was run in a single session lasting approximately one hour. Subjects were tested individually.

After a familiarization phase, subjects had to analyze a source example. In the first step subjects searched for the solution to the example problem presented on a chessboard, i.e., how the white player could put the black king in checkmate in a few moves. None of the subjects found the solution in the allotted time (1 min.). The second step involved explaining to half of the subjects ("Case" condition) the exact solution procedure for this particular example, and to the other half ("Principle" condition), the general principle of smothered mate with sacrifice, illustrated with the example. The subjects then had to reproduce the correct procedure on the chessboard while explaining the moves.

Then, the subjects had to solve two problems, one " like " the example (both in its structural and visual features) and one " unlike " the example (a problem that looked different on the surface but was in fact structuralley isomorphic). The time limit was set at 4 minutes per problem. Whenever the correct solution was found, the solving time was recorded.

After this problem solving phase, the subjects were given an empty chessboard and the complete set of chessmen, and were asked to recall, as fully and accurately as possible, the layout of the example initially explained by the experimenter.

Finally, the example layout was presented to the subject, and he or she was also given three other layouts and asked to order them in decreasing order of similarity (in terms of the required solving steps) to the example layout.

Summary of results

Table 1 summarizes the results. The subjects in the two groups (Principle and Case) were distinguished on the basis of their perfor-

	Principle Group (N = 22)			Case Group (N = 22)		
	(n=7)	(n=10)	(n=5)	(n=1)	(n=9)	(n=12)
Performance profile	+ +	+ -	- -	+ +	+ -	- -
High recall	71%	60%	80%	100%	44%	42%
Reconstruction capability	86%	70%	40%	100%	44%	25%
Structural criteria	86%	80%	80%	100%	67%	42%

Table 1. Summary of results for the two groups of subjects (Principle and Case): performance profiles on like and unlike problems, recall test performance, reconstruction capability, and similarity based on structural criteria.

mance profile on like and unlike problems (+ +, success on the two kinds of problems, + -, success only on the " like " problem, - -, failure on the two problems). In each case, the table gives (i) the percentage of subjects with a "high" score on the recall test (at least four pieces placed in the correct location), (ii) the percentage of subjects capable of reconstruction (they put at least one relevant piece in a logical location that did not change the structure of the game), and (iii) the percentage of subjects whose similarity order placed priority on structure.

These results indicate some very important differences between the two experimental conditions, " Principle " and " Case ". Differences were found not only in the scores on the like and unlike problems, but also on example recall and on judgments of new problem similarity.

Among the subjects who succeeded on both types of problems — all but one of whom belonged to the Principle group — most seemed to remember the example well and were capable of reconstructing the game without changing its structure. In addition, most subjects identified the structure of the new problems and used this criterion to determine how close they were to the example.

For subjects who succeeded on the like problem only or who failed on both problems (both groups contained such subjects), the results showed that there were still large differences between the two conditions on the recall and similarity tasks. Case group subjects who only succeeded on the like problem exhibited poorer performance on the recall and similarity tests than Principle group subjects with the same profile: Case group subjects remembered the example less accurately, were less often capable of reconstruction, and outnumbered the others in relying on surface features to decide how similar the new problems were to the example. The same types of differences between the two groups were observed for subjects who were unable to correctly solve either problem. While the non-solvers in the Principle group remembered the example well and some of them were able to reconstruct the game, the corresponding Case group subjects did not remember the example as well and very few of them could reconstruct. In addition, the non-solvers in the Principle group primarily used a structural criterion for judging new problem similarity, whereas a majority of the non-solving Case subjects relied mainly on surface criteria.

DISCUSSION

This experiment pointed out the existence of different ways of solving problems by analogy: the use of an abstract schema and/or adaptation of a source case. Various measures enabled us to identify different forms of source example processing, storage, and retrieval for solving new isomorphic problems. These experimental results thus support the hypothesis that during learning, representation structures of different levels of abstraction are elaborated, and that access to these different structures depends on the similarity of the to-be-solved target problem to the already-processed source problem.

We devised an experimental setup that led subjects to use different encoding methods to learn the example problem, which involved winning a chess game in a given way. Some subjects were simply shown the remaining steps needed to win in this particular case. Others were given the general solving principle for this type of game ending (smothered mate with sacrifice), which was illustrated using the same example. Learning was assessed by having subjects solve two new problems from the same problem class, one like the example in its surface features and structure and one that was superficially unlike the example but was isomorphic to it from the structural standpoint.

The results showed that some subjects correctly solved both types of target problems, others, only the like problem, and still others, neither problem. In line with our assumption that exposure to an abstract principle promotes learning (e.g., Clement, 1994; Catrambone, 1995, 1996), all subjects who succeeded on both types of problems (except one) were subjects who had been presented with the abstract solving principle. These results are thus compatible with the hypothesis that to be able solve all problems in the problem class studied here, no matter how close the target problems are to the source, it is necessary to construct an abstract solving schema.

However, mere exposure to the abstract principle did not induce a knowledge level in all subjects that enabled them to solve both problems. Many subjects only succeeded on the like problem, and others, on neither problem. Subjects in the group that was only given the specific procedure for solving the example, failed on one or both problems.

Subjects were found to be sensitive to surface similarities between the target problem and the example, even those who succeeded on both new problems. Solving times were shorter for the like problem than for the unlike one. These results support the hypothesized existence of two distinct processes: (1) a search-and-adapt process that searches for the case and adapts it to the new problem, and (2) an apply-and-instantiate process that applies an abstract schema and instantiates it with the specific data from the target problem. Subjects may rely on one or the other process, depending on how similar the target is to the example (e.g., Brooks, Norman, & Allen, 1991, Gobet & Simon, 1996a, 1996b; Pierce et al., 1996). When the problem to be solved is deemed by the subjects to be like an already learned problem, they access that problem and attempt to adapt it to the solution. When the to-be-solved problem differs from the source problem in its surface features, subjects access the abstract schema and attempt to apply it, while taking the specific features of the new problem into account. For subjects who succeeded on one problem only, it was always the like problem. So these subjects must not have built an abstract schema and were thus limited to adapting the solving procedure of the example to the target problem. This was only possible when both the target problem's surface features and structure were very similar to those of the example. For subjects who failed on both problems, adaptation of the example to the target problem must not have been possible, even when the two were very similar (e.g., Reed & Bolstad, 1991).

The data we collected provide further insight into the nature of the representations constructed and used by subjects. Our results clearly showed that subjects who succeeded on the like and unlike problems had acquired general knowledge for solving problems in this class as reflected by (a) the fact that they were able to reconstruct the example without changing its structure, even if the pieces were not placed in their correct locations, and (b) the fact that they were able to assess the similarity of new problems on the basis of a structural criterion. But these subjects were also the ones who were better at remembering the source problem: they stored the relevant features of this specific example in memory. Concerning the subjects who only succeeded on the like problem, our analyses pointed out substantial differences in the way the problems were encoded, depending on whether or not the subjects had benefited from exposure to the abstract solving principle. Those subjects who had been exposed to the general principle usually remembered the example more accurately than the other subjects did; they were also better able to reconstruct the example without changing its structure, and they placed more priority on structure in judging how similar new problems were to the example. Analogous results were obtained for subjects who could not solve either problem. It thus appears as though the mere analysis of the subjects' performance profiles on problems that are like and unlike the example is insufficient for determining how the example was encoded. This brings us to the more general issue of how source problems are encoded when a new class of problems is being learned.

In attempting to define the different encoding modes used in problem solving and determine how they evolve with learning, it would certainly be a gross oversimplification to distinguish only the storage of special cases and the building of abstract schemas. In line with classical theories of memory (see Tulving & Thompson, 1973; Tulving, 1985; for a review see, Tiberghien, 1997), it would no doubt be more useful to hypothesize that there is co-construction and co-existence of different types of problem encoding, some more perceptual in nature, others more episodic and procedural, and still others, more semantic and conceptual.

REFERENCES

Anderson, J.R., Fincham, J.M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 932-945.*

Brooks, L.R., Norman, G.R., & Allen, L.R. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120, 278-287.*

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20,* 493-523.

Catrambone, R. (1995). Aiding subgoal learning: Effects on transfer. *Journal of Educational Psychology, 87,* 5-17.

Catrambone, R. (1996). Generalizing solution procedures learned from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1020-1031.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

Clement, C.A. (1994). Effect of structural embedding on analogical transfer: Manifest versus latent analogs. *American Journal of Psychology, 107,* 1-39.

Cummins (1992). Role of analogical reasoning in the induction of problem categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5,* 1103-1124.

Gick, M., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.

Gobet, F., & Simon, H. A. (1996a). Recall of random and distorted chess positions: implications for the theory of expertise. *Memory & Cognition, 24,* 493-503.

Gobet, F., & Simon, H. A. (1996b). Templates in chess memory: a mechanism for recalling several boards. *Cognitive Psychology, 31,* 1-40.

Pierce, K. A., Crain, R. M., Gholson, B., Smither, D., & Rabinowitz, F. M. (1996). The source of children's errors during nonisomorphic analogical transfer: script theory and structure mapping theory. *Journal of Experimental Child Psychology, 62,* 102-130.

Reed, S. K., & Bolstad, C. A. (1991). Use of examples and procedures in problem solving. *Journal of Experimental Psy-*

chology: Learning, Memory, and Cognition, 17, 753-766.

Tiberghien, G. (1997). *La mèmoire oubliée*, Mardaga.

Tulving, E. How many systems memory are there? *American Psychologist, 40*, 385-398.

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373.

ANALOGICAL CONSTRUCTION OF A SYSTEM OF SIMULATION : A CASE STUDY IN PHYSICS

Sandra Bruno
%G. Vergnaud, Equipe Cognition et Activités Finalisées.
CNRS - Université ParisVIII
2, rue de la liberté 93 526 Saint Denis Cedex 2 - FRANCE
Tel : 33 1 43 62 62 65 e-mail : brunosandra@yahoo.com

INTRODUCTION

In this paper, our attention will be centered around how conceptualization comes into being and can be a part of the analogical reasoning process.

THREE ANALOGICAL SYSTEMS

We will distinguish three types of situations upon which an analogy takes place. This will be done on the criteria of their relational structure.

Analogy between proportions

It is the classical schema "A is to B what C is to D" (figure 1). In this kind of analogy, there is one invariant (explicit or not) that permits a similitude between two pairs of objects.

Analogy between systems

It is an isomorphism between two relational structures (figure 2), as it can be seen in Rutherford's solar system and atom (Gentner, 1983), or in Gick & Holyoak's (1983) fortress attack and radiation problems. Most of the experiments and theories on analogical reasoning are based on this type of situation.

Analogy between modified systems

A new isomorphism is derived from analogous situations, each modified by analogous transformations (figure 3).The new pair of analogous situations can then be considered as a complexification of the initial one : more objects and properties are to be considered. Gentner's (1983) use of the analogy between electricity and flowing water is a good example of analogy between modified systems. However, it is regrettable that this context of reasoning is very seldom exploited in psychological research on A.R, since it is a frequent approach in scientific modeling (think of the A.I metaphor of human cognition, with its basic isomorphism, and the large scope of possible variations).

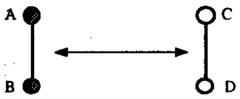


Fig. 1. Analogy between proportions

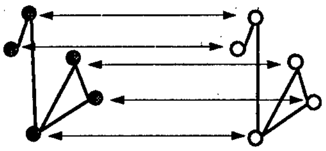


Fig. 2. Analogy between systems

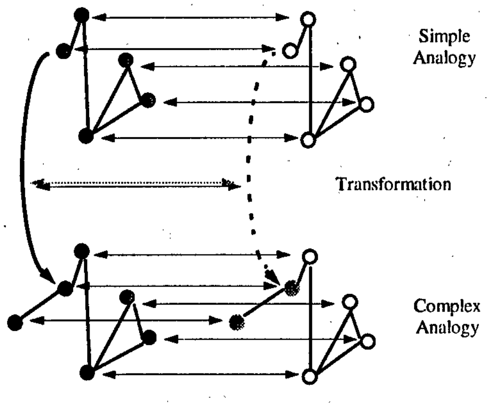


Fig. 3.Analogy between modified systems

CONCEPTUALIZATION IN MODIFIED ANALOGOUS SYSTEMS

We have examined logical aspects of analogical situations ; we will now consider psychological aspects, in particular with regard to the analogy between modified systems, where the question of conceptualization arises in a very acute way.

Issue

We would like to introduce the theoretical issue with one of the results obtained by Gentner and Gentner (1983).

Their goal was to "test the Generative Analogy hypothesis : that conceptual inferences in the target follow predictably from the use of a given base domain as an analogical model. To confirm this hypothesis, it must be shown that the inferences people make in a topic domain vary according to the analogies they use" (p.100). More precisely, among other predictions they thought that giving subjects a hydraulic/electricity analogy would facilitate their inferences when one battery was added in series or in parallel to the electric circuits (thus making two different complex situations) . The authors' reason was that "with reservoirs, the correct inferences for series versus parallel can be derived by keeping track of the resulting height of water". This prediction was not supported by the results (but others were, with other analogies). The authors put forward two possible interpretations for this : the lack of knowledge in the source domain and the failure to notice and use the analogy.

Even if it may not be determinant in this particular context[1] we would like to suggest another reason to illustrate a paradoxical aspect in analogical reasoning.

When introducing to the reader the hydraulic conceptual field as an electric analog, the authors proposed to "consider what happens when two reservoirs are **connected in series, one on top of the other**" (p. 113, stress is mine). The fact is: how can one "guess" that reservoirs in series are placed one on top of the other, but not one behind the other like in the spatial organization of batteries, for example. The reason for the right configuration lies in the correspondence "pressure/voltage" and in the fact that doubling the height of water doubles its pressure, in the same way that two batteries in series double the voltage. Thus, it appears that the solution to the problem is required in order to make the right analogy. This constitutes a paradox as it is generally considered that it is the use of the right analog that triggers the solution to the problem.

Thus, we have sought to know whether this paradox was experienced by the individuals during the analogical reasoning, and if so, how it is dealt with. Tackling this question (which, in our view, has been omitted in psychological research, even if Clement's (1988) work comes close) requires a detailed qualitative approach. This is why we chose, as a first step, to proceed by case studies.

Expertise

The conceptual domains used for the experiment are fluid flow and heat flow. The main characterization of these phenomena is the evolution towards a balance state between a source (S) and a receptor (R), linked together with an intermediary element (I). We will introduce them in a phenomenological way, without using the mathematical relationships usually used to describe the physical laws.

The heat flow

The complex phenomenon to be simulated with a hydraulic setting is the heat flow emanating from a source through any material, a piece of wood for example. Figure 4 represents the setting for this thermal phenomenon [Th2], also showing the heat flow direction, which we know is orientated towards the colder part of the material.

[Th2]

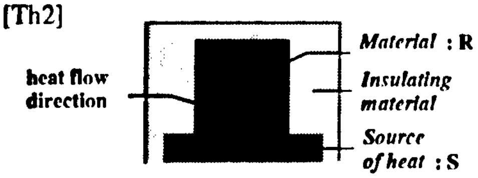


Fig. 4. A complex thermal setting - any material (piece of wood)

---

[1] What the subjects were taught during the training phase of the experiment was not detailed in the article. Were they taught only the simple analogy, or the complex one as well (with more elements in the circuit)?

The objective is to study the spread of heat in the material, using timed measurements of the temperature (T), at different points (ideally an infinite number) in between the source of heat and the top of the material. Thus, the temperature function has one variable : the time (t), and one parameter : the distance from the source (d).

As often done in physics, before studying the overall system we will consider the "smallest" possible part of it : a "slice", which will constitute the simple system. A simple thermal setting ([Th1], figure 5) can be made to experiment on how the heat behaves in a slice of material, by modeling the decomposition of its two appropriate material properties toward the heat flux:

- the conductivity (K, how much heat is passing through the material per unit of time) will be represented by a piece of paper (I) acting as intermediary and whose capacity can be neglected
- the thermal capacity (C, number of heat units required to raise the temperature of the material by one degree) will be represented by a piece of copper (R), acting as a receptor, which can be considered isothermal because of its very high conductivity.

[Th1]

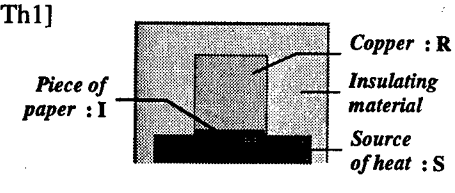


Fig 5.A simple thermal setting (" slice")

A thermometer is placed on top of the piece of copper, in order to measure the evolution of the temperature in function of the time, which remains the variable. Because of the copper's isothermal property, the distance parameter is no longer significant, and is therefore not taken into account.

The hydraulic flow

In a hydraulic setting ([Hy1], figure 6), the liquid flow can simulate the heat flow in a slice (the evolutions will be the same). The source (a large beaker containing a liquid) is connected to a receptor (a thin beaker), via a pipe of very small diameter. The conductivity of such setting is proportional to the diameter and length of the pipe, and to the viscosity of the liquid. The capacity of the receptor is proportional to the diameter of the small beaker.

[Hy1]

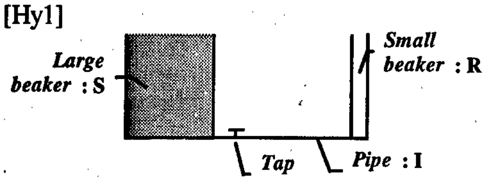
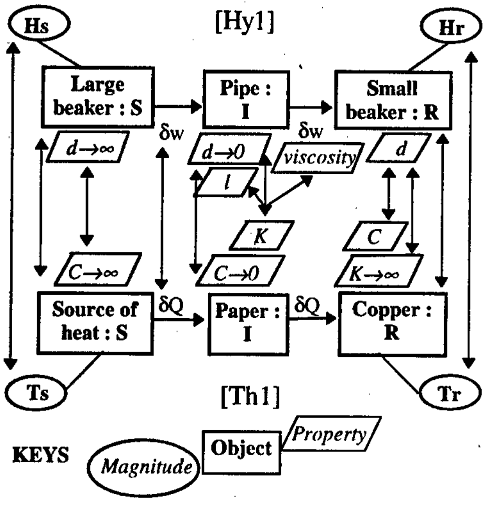


Fig. 6. A simple hydraulic setting (simulating the heat slice)

The analogy

Figure 7 sums up the simple analogy between the simple hydraulic and thermal systems.



H:Height ; w:weight (of liquid) ; d:diameter ; l:length
T:Temperature ; Q:heat ; C:Capacity ; K:conductivity

Fig. 7. The simple analogy (correspondences between objects, properties, magnitudes and flux)

EXPERIMENT

The aim of the experiment was to observe how the subjects construct a complex hydraulic analog, after having been given knowledge of the simple analogy and the complex thermal system to be simulated.

Subjects and method

The 8 subjects taking part in the experiment were first year physics students at university. They attended a 4 hours practical class, initially devised independently of our study. It consisted mainly in making experiments, taking down data and tracing the graphs (this was followed, one week later, by a theoretical class, for the interpretation of the results).

In order to introduce a problem situation, the psychologist and the teacher adapted the pedagogical scenario, to :

1- the teacher introduces the simple experiments [Hy1] and [Th1] as exposed above (§Expertise).

2- the students (in pairs) make the experiments [Hy1] and [Th1].

3- the students are asked to cite all the analogies between [Hy1] and [Th1] they have noticed during the experiments.

4- the teacher exposes the simple relevant analogies (the one exposed above in §The analogy, plus mathematical ones).

5- the teacher introduces the problem of heat propagation in wood, and ask the students "to find a hydraulic analogy that simulates the wood setting and the heat evolution".

6- the students construct the complex analogy.

Data collecting and processing

An observer accompanied the pairs of students. Their conversations were tape recorded, their drawings were collected.

During phases 2, 3 and 6 of the previous pedagogical scenario, the observer could interact with the students.

The transcription of the audiotapes was processed in three steps, leading to :

- sequences : units of discourse (Œ...)
- micro-units : sequences relevant extracts
- reading table : formalized summary of the micro-units (ex : R[Hy1] indicates the receptor of the simple hydraulic setting) .

Results

We are concerned here with phase 6 of the pedagogical scenario. We present a protocol of two subjects.

INTERPRETATION

The interpretation of the protocol will be organized around the three main steps of the subjects' resolution : analysis, conception, improvement[2].

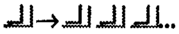
COMPLEX ANALOGIES - Subjects Lea and Gac

Reading table		*Verbalizations*
		●Th1 → Th2 : Analysis of the differences, transformations
	Gac:	*I can't see the difference !*
Th2 : T = f(d,t) Th1 : T = f(t)	Lea:	*Because here* [Th2], *the temperature depends on the height and on the time, and with the copper, as it goes very quickly, it was only the temperature. We neglected X* . (X is the letter given by the students to the distance parameter).
		●Th ↔ Hy : Mapping of the parameters, understanding of the problem
X[Th2] ≡ H[Hy2] T[Th1] ≡ H[Hy1]	Lea:	*It is ... X would become H ... on the other hand, T was in fact H in the experiment ..... The temperature was H, it was the height.*
	Lea:	*Therefore it must depend on two parameters.*
T[Th2] ≡ H[Hy2] X[Th2] ≡ ???	Lea:	*The temperature corresponds to the height, well, we keep the time but X, I don't know what would be its correspondence, hum, for the hydraulic system*
		●Th2 ↔ Hy1 : Comparison of the evolutions and assimilation of the objects
	Obs:	*It* R[Th2] *warms up little by little*
	Lea:	*but this* R[Hy1] *fills up hum...*
f(t) [Th2] ≡ f(t) [Hy1]	Gac:	*little by little also*
⇒ R[Hy1] ≡ R[Th2]	Lea:	*little by little also but ...*
≡ R[Hy2]	Gac:	*In fact, this* R[Hy1], *we consider that it is... the entire piece of wood.*

❹Th1 → Th2 , Th2 → Th1 : Analysis of the transformations

ΣQ[Th1] = Q[Th2]	Gae: *And what we calculated with the temperature* [Th1] *in fact.... was the quantity that we had each time, ticticticic.... well that increased! in relation to the overall volume.. in, well, in the wood analogy* [Th2].
New magnitude : ρ	Gae: *In fact it is the density, it is the density that increases.*
Principle : Division	Gae: *In fact we need to devide ... the big volume by the amount of what arrives each time.*

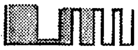
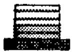
❺Th1 → Th2 ⇒ Hy1 → Hy2 : Application of the Th transformation in the Hy setting

	Lea: *You mean to say that we should .... put loads of little reservoirs ?*
Principle : Repartition	Gae: (disregards Lea's proposal*) Then... there is a little heat that arrives and spreads around*
	Lea: *all right, then we should make loads of little reservoir?*
Verification : reverse the division principle	Lea: *Well* [Hy1] *it is the same as the piece of copper because when .. the water arrives it .. well .. it can't be separated*

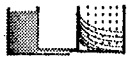
❺bis Th1 → Th2 ⇒ Hy1 → Hy2 : Application of the Th transformation on a Hy principle

$\rho = f(h)$ [Hy2] ≡ $T = f(d)$ [Th2]	Gae: *The water should.. in fact.. each time.. it would be the same level but it is only the density that changes, this would be equivalent to the temperature.*

❻ Hy2 : Setting proposal, justification

Hy2 : 	Lea: *The water would arrive hum in a small reservoir and we need this reservoir to be filled in order for it to give some to the other one*
≡ Th2 : 	Lea: *Because, if we devide the wood in many small parts, a first small part must be heated so that it can give some heat to the other part.*
	Lea : *The heat arrives here, it fills up here first, then it goes to the other one, and here it's OK.*
Objection	Lea: *The problem is that for the wood system it's an infinitesimal quantity.*
Identification param.	Gae: *Therefore we've got at last the magnitude X*

❼ Construction and integration of the propagation law

	Obs: *But is there one that is completely warm before it goes to the other?*
	Gae: *no it wouldn't be exactly like that, it gives a little bit*
	Gae: *And in fact, that is what the analogy is: the water arrives like this, we have a first small beaker .. and Hop! it gives some to the other one ..*
	Lea: *It's filled in through minute holes because it slowly gives out some drops*
$H(d) > H(d+\delta)$ [Hy2] ≡ $T(d) > T(d+\delta)$ [Th2]	Gae: *Therefore it is like for the wood analogy.. we don't need to wait until it fills up completely for the water to go to the other one*
	Gae: *And the more it fills up, the more it gives*
	Lea: *And the front-line of heat is preceded by a little added heat, in fact that's what it is!*

❽Hy2 : Integration of a property : Conductivity, speed factor

	Obs: *And here* [Hy1], *it was a slowing down motion here. ..*
Holes [Hy2]	Lea: *In fact the holes are so small that that's what makes it slow down.*
≡ Pipe [Hy1]	Gae: *No the holes can be bigger, it depends*
	Lea: *It should be more or less proportional to the wood conductivity*
$\delta Q = K(T(d) - T(d+\delta))$	Gae: *Yes you bring in a K factor in fact*
≡	*Lea: Yes, because what would be perfect would be to have a kind of porous material in order to ..*
$\delta V = K(H(d) - H(d+\delta))$	Gae: *The more there is the more will drop and .. a coffee filter, that's very good!*

---

[2] The development of the steps may seem to be an exemplary canonic model! But we underline that this finding became obvious after the cutting out in sequences following the described manner (See §Data precessing).

❶→❹: *Analysis of the systems considered in the task : highlight of the differences, operations of transformation, of mapping, of assimilation*

The subject Lea describes the difference between the simple and the complex thermal systems with regard to magnitudes (temperature taken at one point, or taken depending on the distance from the energy source). She explains this difference by using, in an implicit manner, the properties of the objects ("*the copper, as it goes very quickly...*" points out the very high conductivity). She then looks for a hydraulic correspondence to the additional magnitude of the complex thermal system (the distance), which she will not identify. Two comments on this matter: firstly, from the A.R point of view, the subject understood that the problem was to find a magnitude corresponding to the distance; secondly, a kind of primitive function of spatial apprehension led the subject to think of making, in a first attempt, a correspondence between the distance (in the heat receptor) and the height of the water (in the hydraulic beaker), probably because both are bottom-up directed. For some time, this primitive function is predominant in Gae's thinking ("*in fact, this* [the receptor beaker], *we consider that it is the entire piece of wood*"). This spatial focusing, which can be considered as an obstacle to the correspondences Hy-Th, will play an important role, as Gae will rely on it to elaborate the relationship between the simple and the complex systems : the wood is the result of the concatenation of the first simple concatenation "copper-paper". Lea tried to go beyond this primitive function after taking the contradiction into account: the height of the water cannot be a parameter of a function as well as the measured value. Thanks to this approach, she will be able to materially interpret the relationship established by Gae, as is shown in the next step.

❺→❻: *Proposal for a setting and for the evolution of the phenomenon*

Lea applies the concatenation to use it as a transformation of the simple hydraulic setting towards the complex hydraulic setting ("*we should put loads of little reservoirs*"). It is important to note here that at this moment, the subject doesn't mention how the reservoirs must be linked together. In giving up his spatial apprehension with regard to the distance-height equivalence (where water density should vary in function of its height as the amount of heat varies with the distance!), Gae will agree with Lea's proposal. In order to verify the physical coherence of the proposed setting, and to make precise its configuration, the subjects utilize the supposed evolution of the complex phenomenon in the heat and hydraulic systems ("*the heat arrives here, it fills up here first...*"). They will come to the conclusion that their hydraulic setting answers the question brought forward in the previous task analysis : "*Therefore we've got at last the magnitude X*". However, Lea still shows signs of further hesitation ("*the problem is that, with the wood system, it's an infinitesimal quantity*"), and will not pick up on it herself but she will use this comment for the following improvements.

❼→❽: *Improvement of the setting by integrating constraints, and putting down laws*

A first outline of a complex analogy has been marked, for which a certain number of elements constitute the foundations (in particular the concatenation relation), and others can now be removed or modified.

Without any resistance, the subjects dismiss the physics principle that justified their first setting and modify it, taking into account principles that had not been previously considered ("*we don't need to wait until it fills up completely for the water to go into the other one*" ; "[the holes are] *proportional to the wood conductivity*"). In these extracts, it appears that the material analogy of the settings provides some kind of assistance and some thoughts materialization in order to verify the relevance of the proposed principle (a kind of meta-cognitive formula could be: "in order to validate the hypothesis of this principle, it must be applied for the thermal as well as for the hydraulic setting").

In parallel to this materialization, the subjects, using their spontaneous means of expression, refer to two fundamental laws which encompass the overall phenomenon : the flux (of heat, water) is proportional to the difference between the source and the receptor of the considered magnitude (height, temperature): "*the more it fills up, the more it gives*", and the flux is proportional to a conductivity property which depends on the objects, and which can be measured.

DISCUSSION

1- The analogical reasoning brings into play various cognitive operations

Mapping and transfer are generally the two main cognitive processes used to describe the A.R. We found occurrences of these processes, but they seem to be insufficient to describe analogical modeling.

Also, procedures like the identification of differences (between the simple and complex systems : there is one more magnitude to take into account ; between the complex systems : the first hydraulic setting does not answer the infinitesimal characteristic of the wood system), procedures like systems transformations (in particular by means of the concatenation relation, but also more generally the transformations authorized by the objects taken into account and their properties), and procedures like assimilation, play a role that should not be disregarded. In the last-mentioned procedure, an element is extracted from the context of the system to which it belongs and imported into the system in focus (Gae "sees" the receptor beaker as the piece of wood ; Lea assimilates the heat flux to her complex hydraulic system, using terms of the hydraulic domain (*"the heat (..) fills up"*)).

Additionally, transfer manifests itself in our protocol like a mechanism that is not unidirectional. Evolutions, relations and laws are imported from one domain to the other and vice-versa. For example, the subjects realize that the heat propagation is not "in stairway" as it is the case for the water propagation in the first complex system proposed. They then import the principle of "progressive propagation" into the hydraulic setting and transform the setting for the principle to be applied. In return, the subjects will import into the thermal system the law of speed propagation understood in the new hydraulic system, probably thanks to the particularly visual aspect of the evolution of the hydraulic phenomenon.

Lastly, these comings and goings, together with assimilation operations, seem to be cognitively important for the construction of conceptual and relational invariants.

2- The analogical reasoning implies representations of various conceptual registers.

We observe the presence of various conceptual registers : concepts related to the material objects of the setting (mainly the kind and configuration of the receptor) ; the objects properties (conductivity, diameter of the holes) ; the magnitudes representative of the evolution of the phenomena (the temperature, the water level) ; relationships between objects (mainly the concatenation), and between magnitudes (functions, laws).

We suggest the hypothesis that, in search of a better coherence in the current stage of his reasoning, the subject confronts these various registers, aiming at the re-adjustment of the activated knowledge.

In the A.R, this confrontation could be the driving force behind these comings and goings between the systems, each presenting different local and temporary facilities.

3- The analogical reasoning within a modeling task can generate learnings by way of awareness processes.

In some ways, the laws cited by the subjects were not unknown to them, as they were part of the informations delivered by the teacher.

However, it is clear that these laws take another dimension at the outcome of the simulation work, and of the awareness triggered by this work.

It is indeed symptomatic to observe that these laws, which form the starting point of an expert modeling work, are cited by the students at the end of the protocol.

Furthermore, it is also symptomatic to observe that these laws were not expressed by the students in a formal or canonized manner; it is probable that at that precise moment, these "informal laws" have not precisely the status of a law for them... but they are ready to receive further explicitation from the teacher.

CONCLUSION

1- The paradox of analogical reasoning

With concern to our initial theoretical issue, the points developed in the discussion throw some light on the way subjects manage the paradoxical aspect of A.R, in which they need to anticipate "what's going on" to construct a physical setting, and at the same time need to construct a setting to understand "what's going on". It seems that these two sides of the paradox are inherent to the analogical reasoning, and even may support the conceptualizations of the phenomena. Thus, the subjects would construct physics laws in order to test the coherence "setting/evolution", by creating a relationship between the properties of the elements, and the flux "authorized" by the properties and the configuration of these elements.

2- The study of analogical reasoning and a theory of representations

The elements of the discussion together with the preceding conclusion lead, in our view, to the idea that the study of analogical reasoning must be included into a theory of representations, like the one of the homomorphism "real-representation" developed by Vergnaud (1987). Indurkhya (1992) modeled the "similarity creating metaphors" in reference to Holland's model, which also postulate this homomorphism. But, in the one hand metaphor and analogy are rather different processes, and on the other hand, Indurkhya's work on A.R takes into consideration little psychological data.

Our perspective is to bring more elements in this direction.

REFERENCES

Clement, J. (1988). Observed Methods for Generating Analogies in Scientific Problem Solving. *Cognitive Science*, 12, 563-586.

Gentner, D. & Gentner, D. (1983). Flowing Water or Teeming Crowds : Mental Models of Electricity. In D. Gentner & A. Stevens (Eds.), *Mental Models*. Hillsdale, NJ : Lawrence Erlbaum.

Gick, M. & Holyoak, K. (1983). Schema Induction and Analogical Transfer. *Cognitive Psychology*, 15, 1-38.

Indurkhya, B. (1992). *Metaphor and Cognition*. Kluwer Academic Publishers.

Vergnaud, G. (1987). Les fonctions de l'action et de la symbolisation dans la formation des connaissances chez l'enfant. In J. Piaget, P. Mounoud & J.P. Bronckart, (Eds.), *Psychologie, Encyclopédie la Pléiade*, XLVI (pp. 821-844). Paris: Gallimard.

Special thanks to P. Mathieu and S. Nadot for their active participation to the experiments, and to S. Mordacq for the translation.

THE COPYCAT PROJECT: TOWARDS A CONCEPTUAL FLUIDITY THEORY

Bruno Vivicorsi
CREPCO UMR 6561
(Center for Research in Cognitive Psychology)
Université de Provence & CNRS
29, av. R. Schuman, 13621 Aix-en-Provence Cedex 1, France
vivicors@newsup.univ-mrs.fr

INTRODUCTION:THE MENTAL FLUIDITY

Mental fluidity (or conceptual adaptation) appears in a lot ofactivities that are more or less general, like analogy-making, understanding metaphors or puns, translating or contracting texts, imagining tobe another person, counter-factuals-making, human language error-making,humor-making, music-playing in another style, words-blending,forms-recognising, conceptual learning, rolegames-playing, publicity-understanding, science-fiction, politics, poetry— and this list is not exhaustive... All of these activitiesrequire the use of analogies.

We generally think that an analogy is when the subject finds the bestmatching between elements of two analogous situations, but there is alsothe fact that we perceive situations in a certain way and *then* make correspondences between some elements of these situations. We notknow all that we should know to act on the world, but we know all we shouldknow when we solve a problem in an experiment in which there is all thenecessary informations, and then we neglict a part of the analogical process, aperceptual part (Chalmers *et al.*, 1992).

The capacity to perceive analogies between two objects, situations orfacts at a certain level of abstraction is the general capacity that appears in activities that require aconceptual fluidity. This capacity is very natural: we can conceptuallyadapt ourselves at any new situation without research in a listing whathappens, and without try to understand what changes in comparison of the time before. Indeed, we immediately see*what is the same* — at a certain level ofabstraction (*i.e.*, to see a thing *as*another thing depending on pressures) — and try to use these informations to respond to the situation's problem.Consequently, the nature of analogy-making is seen here as a generalcognitive process rather than an exceptional mechanism brought to bear only in unusual circumstances, and the resolution of an analogy is seen as atranslation (or adaptation) from a structure to another (Hofstadter,1985).

THE MECHANISMS: THE COPYCAT MODEL

The copycat's microworld

A microworld of letters was created by Douglas Hofstadter tostudy more rigorously the mechanisms of the mental fluidity (see Hofstadter*et al.*, 1995). The microworld is composed by the alphabetletters andis a no circular alphabet, each letter having a knowledge of his neighborletter. The COPYCAT project (Hofstadter & Mitchell, 1994), based on thisletter-strings microworld, illustrates this process with creative analogyproblems as «suppose the letter-string **abc** were changed to **abd**,how would you change the letter-string **mrrjjj** in "thesame way"?[1] ». This world reduction is necessary to grasp in a clear way all theoperations that are used between the perception of a creative analogyproblem and the problem's resolution. The analogies are creative in thesense that the prob-

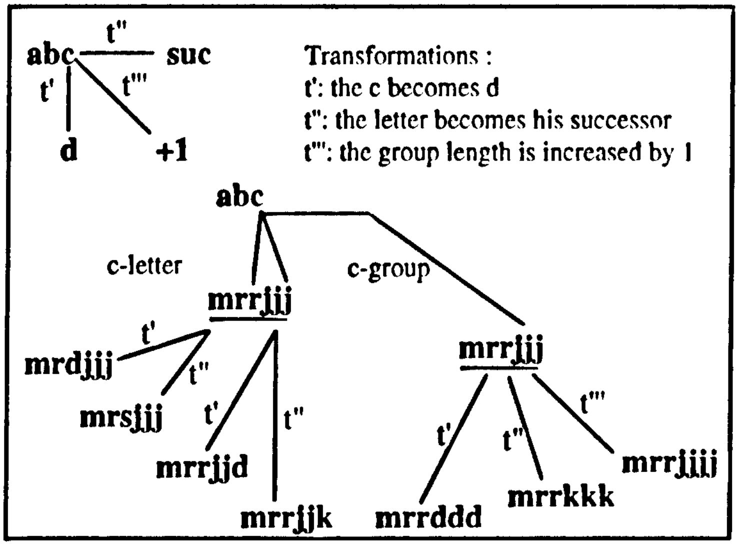


Figure 1. Some solutions to the«abc >> abd, mrrjjj>> ?» problem, depending on the perceivedtransformation and the element of mrrjjj on which the transformation is applied.

lem can have more than one coherent solution, depending on the perception of the transformationbetween the first string and the second string of the problem (Figure1).

What is happened in the transformation of **abc** in **abd**? The **c** is changing? The *third* letter is changing? The *last* letter ischanging? The *higher* letter of the alphabet is changing? And then, what is the element of thestring **mrrjjj** which corresponds to the **c**? Is the *last* letter **j**? The *third* letter **r**? The *last group* of letter **jjj**? But what is exactly the transformation: is the **c** (orthird letter, or last letter) becomes a **d**, or becomes a*successor* of **c**, or a *successor* of the last or higherletter? All these considerations lead to some solutions, but the givensolution depends on the *perception* of the problem —for example, "the last letter becomes**d**", that can lead tothe solution **mrrjjd**. One other response is **mrrjjk**, by matching the **abc**'s letter **c** and the **mrrjjj**'s last letter**j**, and then applying the (perceived)rule "replace last letter by successor". Another higher abstract level of perception is to consider neither theletters nor the letter groups (for example, giving **mrrkkk**), but the group length: perceiving**mrrjjj** as thelength-string **1**(m)-**2**(r)-**3**(j) leads to lenght term response**1**(m)-**2**(r)-**4**(j), so toletter-string **mrrjjjj** with the rule "replace length of last group by successor". The COPYCAT model(Mitchell, 1993) is able to give a solution to a problem depending on theperceived relations and to give another solution to the same problem if theperception of the relationsthe program "have" is different at the beginning of theresolution. The recent extension METACAT developing by Marshall (1997)seems to permit the creation of rich representations of the analogies madein this microworld.

The architecture of the model is based on an interaction of a largenumber of perceptual agents with an associative, overlapping, andcontext-sensitive network of concepts. This particular *anthill-architecture* (Vivicorsi, 1996a) permits the emergence of a robust high-level behaviorfrom the interactions of a great number of low-level nondeterministicperceptual microagents. All the decisions are probabilistic decisions, soat every time the system can move towards a solution or another —depending on all

---

[1] We note from now on a problem like this: «abc >>abd, mrrjjj >> ?».

the perceived or constructed features that are more orless leading towards a specific solution, with no *determined* solution. This probabilistic dynamic provide to the model the capacity, bythe number of possible solutions given to the same problem, to appear moreflexible that the majority of cognitive models.

THE MECHANISMS

Two mechanisms are proposed and implemented in the model togive an account of the flexibility of the COPYCAT program. First, a*high-level perception* mechanism is used to give an account of the encoding in a certain way ofthe problem: we perceive, for example, that it's the third letter that ischanging in «**abc** >> **abd**», and we do then the adapted transformation on «**mrrjjj**>> ?» to lead to the solutions **mrdjjj**or **mrsjjj**. Second, a*perception-conceptualization loop* is necessary to better adapt ourselves to situations, that it is to notseparate the perception of the problem and the cognitive implications forthe resolution of a problem: for example, perceiving the groups **rr** and **jjj** in the string **mrrjjj** entails to conceive **m** as a group; conceiving **m** as a group of oneletter entails to perceive **rr** as a group of two and **jjj** as a group of three; perceiving **jjj** as a group of threeletters entails to conceive **4**(j) as a successor of**3**(j). And this loop can be generalised from one problem to another one, byimmediately perceiving the group **iiii** as **4**(i) — and not, for example, as two groups **ii** — in «**abc** >> **abd**, **iiii** >> ?».

These two mechanisms seem not to be two specific micro-worldmechanisms that only appear in the COPYCAT's world.

Consider the following simple example: a train goes from A point to Bpoint, distant of 60 kilometers, at 30 km/h speed. A fly goes from B to thetrain and when it touches the train, goes back to B , and then goes to thetrain, and return to B, etc., at 120 km/h speed. How many kilometers the fly has covered when thetrain arrives to B? If we perceive the problem as a *distance* problem, the arithmetical operation to find the solution is verydifficult, but if we perceive the problem as a *time* problem, the solution is evident: the train arrives at the B point in 2 hours, so the fly has covered 120 x 2 = 240 kilometers. We don't need asystem that would find the good representation, but we must have the possibility of having some responses influencedby context and concepts, and we must have an interaction between theconstruction representation process and manipulation of theserepresentations. *To perceive a thing in a certain way* is something that we useeveryday: "this dog is a caretaker", for example, isnot an extraordinary thought that require a high level of reflexion (thequestion "Why?" demands this, but we have *already* perceive the dog*as* a caretaker).

In the same way, observing a painting make us to think to somethingelse that is not in the painting, but this thought can make us to perceivethe painting or a part of this in an other way. Thisperception-conceptualization loop is *the link between perception and cognition*, ignored in numerous of psychological theories and in a certain artificialintelligence conception of the cognitive modelisation. For example, theSTRUCTURE MAPPING ENGINE for the analogical reasoning (Gentner, 1989)separates the knowledge of the mapping "engine", and introduces a certain format of representations that permit to the ENGINE to operate in the whishing sense. The problem is raised by-Hofstadter (1995): how representations are formed? How informations areselected? How informations are organized? How can we explain the select of informations that are notconstructed before the mapping? The problem will be raised while perceptionand cognition are viewed as two independent modules (Forbus *etal.*, in press). Indeed, two"modules" can be studied separately without theexistence of the two modules (this can be seen as an large high-levelperception effect: one aspect is seen, and after the other one) but withthe *created hole* between modules that has to be explained[2].

---

[2] This problem is ageneral cognitive science problem: *the level problem.* See Vivicorsi (accepted) for a study of the Fodor's solution (that has tobe rejected) and of the Hofstadter's solution (that has to beconsidered).

Another very instructive example is the BACON model (Langley *et al.*, 1987) as a model of scientific discovery. It is able to discover theKepler's third law of planets movement, but it only has the relevant onesused for derive this law (the average distances between the planet and the sun and their period). So, the system makes a selection before it has toderive the law, but does not give the solution in at less 2 years likeKepler[3] — the students tested do this in one hour, because they are able tofind the good solution with the good informations required to find it(Chalmers *et al.*, 1992). Where is the derivation of the law? There isn't any need ofinformation selection, high-level perception and interactions between whatis perceived and what is conceived with such knowledge apparatus.

Finally, when we categorize objects to make a distinction in, say,three parts,we can place the objects in three different boxes in front of anexperimentalist; but do we this in the quotidian life when we are not in alaboratory with three boxes to fill? The same question could be posed toall experiments in which the attending solutions are a good one and a bad one: to be obliged to respond within astrict scale of solutions is possible and is not in contradiction with thementaI fluidity, but this kind of experiment cannot show the use ofconceptual adaptation.

In conclusion, these mechanisms seem to be involved in all activitiesrequiring a conceptual fluidity, and are clearly defined in a microworld(Hofstadter *et al.*, 1995) with more than one altenativerepresentation (as in the simple"train" example). The *conceptualslippages* (as in the "dog is a caretaker"example) are made on letter, group, same, opposite, etc. concepts (see Mitchell, 1993, for all details) and are explicitely implemented bythe dynamic of the COPYCAT Slipnet (the program concepts network). If thesemechanisms are required for the conceptual fluidity, we must change ourconception of "a concept" to permit to concepts to be integrated in (Vivicorsi,1997). So, the question is: are they psychologically plausible in all"real" activities like the activities mentioned at thebeginning of the paper?

THEIR PSYCHOLOGICAL RELEVANCE: THE COPSYCAT PROJECT

The COPSYCAT project (Vivicorsi, inpreparation) is the examination of the psychological plausibility of the mechanisms postulated in the COPYCATmodel to give a psychological account of the conceptual fluidity appearingin numerous activities. The first experiments (Vivicorsi, 1996b) show thatthe microworld material used by subjects is a real material that can exhibit the mental fluidity of subjectson this microworld. The material used to produce creative analogy problemspermit more than one solution, so it permit to study which solution isproduced by subject and which perception permit to produce it. The ongoing experiment presented here showsthe reality of the high-level perception on this material.

EXPERIMENT

Forty five Université de Provence undergraduates took part in the experiment. For each of them,10 problems have to be resolved, and for each of the problem, 12 solutionshave to be evaluated (Figure 2), with computer presentation. Clearly, thesubject isgiven a problem, gives a solution with no time limit, and evaluates one byone 12 solutions for the running problem (maybe his one) with no timelimit, clicking on "True" — *i.e.*, it's a possible solution to the problem —, or on "False"— *i.e.*, it is not an acceptablesolution. The "True" and "False" propositions varied between 4 and 8 for each problems, so 50% of the twotypes if we consider all the problems. In sum, 60 TT and 60 FF propositionsare presented to each subject. We suppose that all subjects have the same alphabet knowledge (it's a positive aspect of a *world* reduction without the negative aspect of a*subject behavior* reduction).

Three factors are manipulated and each subject is in one on eightconditions: the problem can be presented before each evaluation or not (P);threeexamples can be presented or not (E); problems can be ordained (like inFigure 2) or not (O). All the proposed solutions are randomised for all thesubjects. Consequently, the design is $S < P_2 * E_2 * O_2 >$. There is five subjects for all conditions but in P-notE-notO (n=6) and in notP-E-O (n=9). We will come later on this problem ofsubjects, but remind you that this work is in progress.

---

[3] 13 years according to Chalmers *et al.* (1992).

We register the solution's subject to each problem, the time for eachproposition's evaluation, the type of evaluation (T/F) in comparison withthe correct evaluation (T/F), the order of appearance of each propositionand problem, and the average time response of the subject.

We use to organize data the *Signal Detection Theory* (SDT) (Green & Swets, 1974) inwhich it is possible to analyze in detail the proportion of the fourpossible cases (Figure 3).

This frame of analyze permit to use two indices: the *discrimination index* (d') and the *decision index*(b) (Figure 4). According to this model, a subject's ability to discriminatebetween true items and false items is given by d', the distance between themeans of the true and false distributions in units of the common standarddeviation. The b criterion measures the subject's criterion of decision, that it is:does he prefer raise the risk to miss hits (*i.e.*, torespond F to a T proposition) or to be directed towards false alarm (*i.e.*, to respond T to a F proposition)? The case in which b = 1corresponds to chance decision.

These two indices, on which means can be calculated without neglict some ofthe global variations, are obtained by measuring the proportions of hits (*i.e.*, TT) on the total of T (60) and the proportions offalse alarm (*i.e.*, TF) on the total of F (60)[4].

		Propositions	
		T	F
Responses	T	hits	false alarm
	F	miss	correct reject.
		60	60

Figure 3. The adapted SDT stimulus-response matrix.

PROBLEMS	SOLUTIONS "TRUE"	SOLUTIONS"FALSE"
lmn >> lmo,kji >> ?	kji kjo kjj kjh lji	ijk kij lkj jkl blo xwf kjk
ijk >> ijl, lmfgop >> ?	lmfgoplmfgol lmfgoq lmfgpq lmfgqr lnfhoq	lmfgqq ijlgoq nohiqr kmfgop lmefoplmfgoz
abc >>abd, abbccd >> ?	abbddd abbcce abbcde abbcef	aababe aaaaad aababx uububc aahahc abbccd babcbd aacacd
aabc >> aabd, ijkk >> ?	djkk jjkk hjkk ikkk ijkd ijkl ijdd ijll	iijl iijd jkkk ijlk
abcd>>abcde, mlkji >>?	mlkjie mlkjij mlkji mlkjih nmlkji mlkj lkji	abcdi mlkjii fghijmljjk abcd
abcm>> abcn, rijk >>?	rnnn rijn rijl rjkl nijk sijk	rikl rhij rijh mijk stuv abcs
rst >> rsu, mrrjjj >> ?	mrrjjj mrsjjj mrrjju mrrjjkmrruuu mrrkkk nrsjjk mrrjjjj	mrrklm mrrppp mrrjjf orrjjj
mrs >> mrt, iiii >>?	mrt iiit iiij iitt iijj jjjj iiiii	ijkl mrsstttt iiim mri
ooe >> o,riippp >> ?	r i p ip ipp	riprr pp iippp ppp ii rrip
eqe>> qeq, aaabccc >>?	qeq bbbacbbb baaacccb abbbc	cccbaaa cacbcac abc bbbabbb qqqeppp abcbaqaaaecccq bacb

Figure2. The problems and the proposed solutions.

---

[4] The three examples proposed in four conditions are not considered, buthaving some examples before the test can influence the responses and theevaluations (see next section).

HYPOTHESES, RESULTS AND DISCUSSION

We are working on the two mechanisms supposed to be involvedin the mental fluidity process.

The high-level perception

Hypothesis is that subjects don't perceive all thepossible solutions for a problem, that it is *they don't perceiveall the relations* that permit to produce these responses. The diagram (Figure 5) shows the position of allsubjects (S) within the space on which there is the*machine-behavior* (M) and the *chance-behavior* (C) tables. M representes a subject who makes noerror (d' —> _; b = 1). Crepresents a subject who responds with no criterion of decision (d' = 0,b = 1). The two b's are the same because the two distributions in eachcase are symetric: in the case of M, thedistributions are very distant and in the case of C they are astounded. S represents the set of the 45subjects (d' = 1,876 [s = 1,205], b = 0,542 [s =0,351]). We shows a table of a subject as an illustration (d' = 1,895,b = 0,532).

These global results show that there is a selection of Truepropositions that is not achance selection. Moreover, the possiblestrategy which consists to give a response within the resolution phase, and then wait for the presentation of this oneto recognise it as True is rarely observed (the pattern would correspond tod' > 3, b < 0,2). The d' index is high enough to say that the two distributions arewell differentiated. The b < 1 shows that subjects haverather judged the "true-lity" of the propositions.

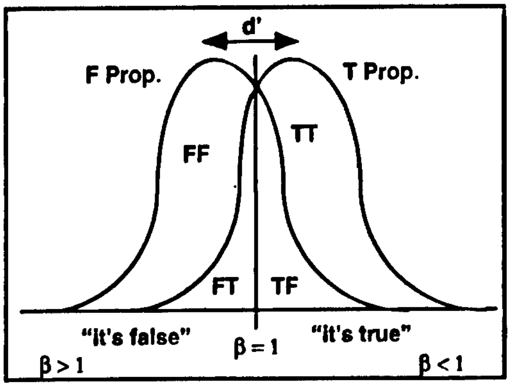


Figure 4. Illustration of the two SDTindices d' and b.

Then, we can conclude of the existence of the high-level perception onhis material, as it was defined in the preview section. But we must gofarther to isolate the strategies (if any) of subjects and to see if there is a different strategy froma condition to an other. The results by condition shows only that in the condition P-notE-notO, there is an *inclination* to adopt the strategy mentioned before. We need then more subjects in eachcondition to analyze the results on which calculating means meanssomething.

Another important indice can permit us to be more precise about the natureof the propositions judged True. Indeed, some propositions are seen as True, but which are produced by the subject before the evaluation? Thehigh-level perception predicts that some solutions are judged True, not all the possible solutions. How manysolutions among the True evaluated ones are perceived in the resolutionphase? Our measure is the comparison with the mean response time of thesubject. The hypothesis is that the subject takes a little time to evaluate the proposition as True if this one was his response for the problem. On all subjects, 70% ofthe subjects responses judged True are given with a time lower that themean response time of the subject. We can then selected what are the solutions activated by subjects before theevaluation test (this work is in progress).

THE PERCEPTION-CONCEPTUALIZATIONLOOP

The hypothesis is that subjects responses or evaluations can influenceother responses and evaluations. We must for this analyze to reorganize allthe patterns with respect of the appearance order, and determine theimplication of one response on the following one. This work is also in progress — we will use the BayesianImplicative Analysis (Bernard & Charron, 1996) for the data treatment.

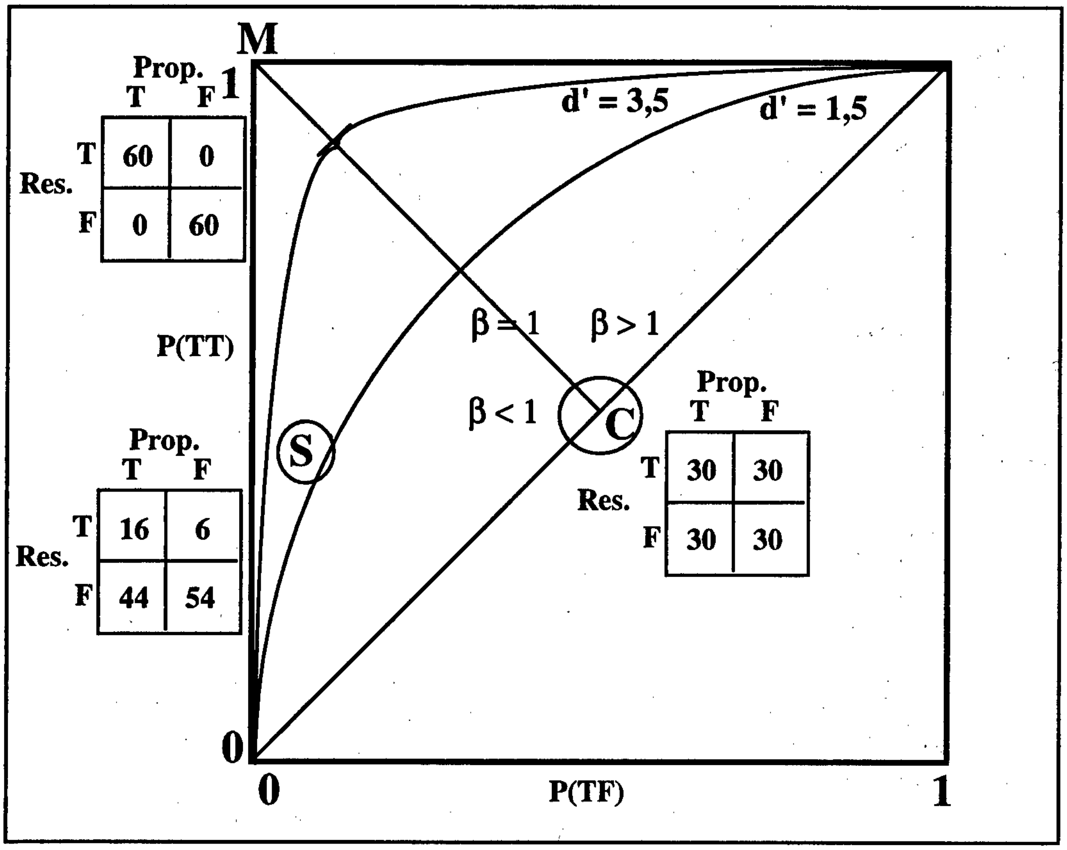


Figure 5. Global results with an example ofa chance-behavior (C), an example of a subject-behavioras an illustration of the set of subjects (S) and theno-error machine-behavior (M).

CONCLUSION

The mental fluidity exists, and we must take it in account for ourresearches, even if it is not necessary that a conceptual adaptation has toappear. The central point is that in a psychological theory or model, aconceptual adaptation *could* appear. We try, with the COPSYCAT project, to evaluate thepsychological relevance of two mechanisms proposed by Hofstadter and Mitchell (see Hofstadter *et al.*, 1995) in the COPYCAT project. These mechanisms are seen here asmechanisms that can give an account of the subjects behavior when they are confronted tocreative problems. The challenge is to determine what is the generality ofthese mechanisms on a more complex world, without reducting the subject's behavior.

This type of research has two important consequences. First, we haveto (re)define the *concept* and *category*terms — indeed, concepts must be *fluids* to be integrated in the mechanisms. Second, the cognitive modelisationmust be constrainted by the perception-cognition loop — *the question is not how many "concepts" are activated,but why these ones are.* The access to a conceptual fluidity theory is difficult, but we must notignore a large part of our activities in order to grasp our naturaltendency to slip from a (micro)world to another (micro)world.

ACKNOWLEDGEMENTS

I would thank F. Antoni, C. Bastien, L. Bourrelly, A. El Ahmadi and A. Tête for helpful experimental and epistemological discussions.

REFERENCES

Bernard, J.-M. & Charron, C. (1996). L'AnalyseImplicative Bayésienne, une méthode pour l'étude desdépendances orientées. I : données binaires. *Math. Inf.Sci. hum.*, *134*, 5-38. II : modèle logique sur un tableau de contingence. *Math. Inf. Sci. hum.*, *135*, 5-18.

Chalmers, D.J., French, R.M. & Hofstadter, D.R. (1992). High-levelperception, representation, and analogy: A critique of artificialintelligence methodology. *J. Expt. Theor. Artif. Intell.*,*4*, 185-211.

Forbus, K.D., Gentner, D., Markman, A.B. & Ferguson, R.W. (in press).Analogy just looks like high level perception: Why a domain-generalapproach to analogical mapping is right. *J. Expt. Theor. Artif.Intell.*

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou& A. Ortony (Eds.), *Similarity and analogical reasoning*(pp.199-241). Cambridge University Press.

Green, D.M. & Swets, J.A. (1974). *Signal detection theory andpsychophysics*. Huntington, NY: Krieger.

Hofstadter, D. (1985). *Metamagical Themas: Questing for theEssence of Mind and Pattern*. New York: Basic Books.

Hofstadter, D.R. (1995). The Ineradicable Eliza Effect and Its Dangers. InD.R. Hofstadter *et al.* (1995, pp.155-168).

Hofstadter, D.R. & The FARG (1995). *Fluid Concepts and CreativeAnalogies. Computer Models of the Fundamental Mechanisms of Thought*. Somerset: The Penguin Press, 1997.

Hofstadter, D.R. & Mitchell, M. (1994). The Copycat Project: A Model ofMental Fluidity and Analogy-Making. In K.J. Holyoak & J.A. Barnden (Eds.),*Advances in Connectionist and Neural Computation Theory, Vol.2:Analogical Connections* (pp.31-112). NJ: Ablex Publishing Corporation.

Langley, P., Simon, H.A., Bradshaw, G.L. & Zytkow, J.M. (1987). *Scientific Discovery*. Cambridge, MA: The MIT Press.

Marshall, J.B. (1997). From Copycat to Metacat: Developing a Self-WatchingFramework for Analogy-Making. In *Proceedings of the Mind IIConferences: Computational Models of Creative Cognition*. Dublin,sept. 15th-17th, Ireland.

Mitchell, M. (1993). *Analogy-Making as Perception: A ComputerModel*. Cambridge, MA: The MIT Press / A Bradford Book.

Vivicorsi, B. (1996a). Les glissements conceptuels, oul'«architecture-fourmilière» comme architecture cognitive. In F.Anceaux & J.-M. Coquery (Eds.), *Actes du 6ᵉ Colloque de l'ARC* (pp.291-295). Villeneuve d'Ascq, dec. 10th-12th, France.

Vivicorsi, B. (1996b). Analogies créatives et architecture cognitive :approche expérimentale du modèle Copycat. Post-Master Degree inCognitive Psychology, Université de Provence, France.

Vivicorsi, B. (1997). "Dog" is not a concept. *Fifth European Congress of Psychology*. Dublin, july 6th-11th,Ireland.

Vivicorsi, B. (accepted). De Fodor à Hofstadter, ou d'un mystère àla nécessité de son dévoilement. *Bulletin dePsychologie*.

Vivicorsi, B. (in preparation). Glissements conceptuels et fluiditémentale : de Copycat à Copsycat. Doctoral Dissertation, Université deProvence, France.

REASONING BY ANALOGY: REPRESENTATION OF PRAGMATIC INFORMATION FROM TARGET KNOWLEDGE INFLUENCES MAPPING

David Cayol
CREPCO UMR6561
(Center for Research in Cognitive Psychology)
Université de Provence & CNRS 29, av. R. Schuman, 13621 Aix-en-Provence CEDEX 1, France
cayol@newsup.univ-mrs.fr

Although "Reasoning by analogy" is an uncommon term for most people, analogical reasoning emerges invarious situations. It is involved in problem solving (Cauzinille-Marmeche,1990 ; Holyoak, Junn, & Billman, 1984), in explanation, in scientific discovery, in creative thinking and so on.

Researches on analogy have been conducted in various domains such as Artificial Intelligence, Neural science, and Psychology fortwenty years. These different approaches have gathered a lot of data which is useful to understand cognitive processes underlying analogical reasoning.

The aim of this paper is to introduce the research about analogical mapping process we have begun during my Ph D. First, Ibriefly outline what is analogical reasoning. Second, SME and ACME models will be expounded. Finally, I will set out the research itself.

ANALOGICAL REASONING

Reasoning by analogy consists in retrieving previous knowledge in order to understand what is unknown or what is new.Authors agree with the idea that reasoning by analogy plays an important role in knowledge acquisition.

It is also possible to characterize this reasoning by its different subprocesses. Subprocesses are representation, retrieval, mapping, transferand induction (Keane, Ledgeway, & Duff, 1994). In order to solve a problemby analogy one must first *represent* the new situation (target problem) and then*retrieve* a useful analogous situation (source or baseproblem).

A core subprocess in analogy is *mapping*. Mapping is necessary for finding out if target and basesituations (or problems) are analogous.This implies that one must construct coherent one-to-one correspondences between two situations. If target and base situations are analogous, *transfering* elements of knowledge from one situation to another is relevant. A classical exemple explaining how mapping progresses is the analogy between the structure of the atom and the structure ofthe solar system (Gentner & Landers, 1985 ; see also Gentner, Rattermann, & Forbus, 1993; Holyoak, & Koh, 1987). The transfer of a portion of the conceptual structure constitutes the basis of analogical *inferences*.

According to Ripoll (1993, 1992), and unlikeour sequential presentation, these 5 subprocesses would concurrentlyrun.

As we have mentioned above, analogical reasoning appears in various usual activities. In addition, Cognitive Psychology has been studying analogical reasoning for about twelve years. Researches have been carried out in Developmental Psychology, Cognitive Psychology, and in Artificial Intelligence.

In Artificial Intelligence, analogy is usefulin two purposes. First, for researchers who aim to understand how the brain functions, analogy is an interesting "mechanism". Second, analogy may constitute heuristic tool toimprove performances of expert systems (Savelli, 1993). Certain systems were elaborated in order to simulate fundamental analogical processes (Gineste, 1997) and in order to investigate how expert systems acquire new knowledge (Cauzinille-Marmeche, Mathieu, & Weil-Barais, 1985).

In Psychology, Piaget proposed a structural stagemodel of analogical reasoning. Piaget and his colleages argued that ability to reason by analogy emerges in early adolescence (Piaget, Montangero, & Billeter, 1977). Accordingly, children would not be able to solve classical analogy task (a : b :: c : d) before being 12 years old, since they could not process abstract relations.

More recently, studies have provided evidencesin favor of the notion that analogical reasoning can be used earlier than the formal operational period (Goswami, 1992;Goswami & Brown, 1990; Holyoak, Junn, & Billman, 1984). These authors have shown that when children understand relations which underly classical (a :b :: c : d) analogies, they manage to complete 4 terms analogy successfully.

We agree with this point of view: we havecarried out a work about analogical problem-solving with young children (5 to 6 years) which has contributed to specifying encodingcircumstances thatfacilitate retrieval process of an analogous base problem (Bastien-Toniazzo, Blaye, & Cayol,1997).

THEORITICAL BACKGROUND

My interest has been turned towards "the core" ofanalogical reasoning: *mapping*. The opinion about analogical mapping we support has become integrated into researches performed by Bastien and Bastien-Toniazzo about context dependence of knowledge.

Bastien argues that knowledge organization is"functional", which means that knowledge is structured with respect togoals to reach (Bastien, 1997).

If analogical reasoning is goal-directed-process (Richard, 1990), like understanding, reasoning and judgment, weassume that analogical mapping process is also goal-directed.

Goal is a context feature in which one acts, one thinks. Activities like reading, understanding, evaluating and problem-solving progress according to goal representation included in current situation. Accordingly, we assess that it is possible to associate the concept of"goal" with the concept of "internal context" (i.e., mental context) proposed by Bastien (1197), because "goal" is included in the representation of new situations.

Mapping models

First models of analogical mapping have attached lot of importance to relational structure.

Two famous models have aimed to simulate analogical mapping. These model sare the Structure Mapping Engine (SME; Gentner 1983; Falkenhainer, Forbus,& Gentner, 1986, 1989) and the Analogical Constraint Mapping Engine (ACME;Holyoak & Thagard, 1989).

Structure Mapping Engine

SME has been elaborated to simulate mapping (M)process: objects (o) from the base (b) knowledge (e.g., thesolar system) are placed in correspondance with objects (o) from the target (t) knowledge (e.g., the structure of the atom):

$M : b_o \rightarrow t_o$

Mapping process is assumed to be governed by "*Systematicity* principle" that plays significant part in SME. Systematicity principle "is a structural expression of our tacit preference for coherence and deductive power in interpreting analogy" (Gentner, 1988, p. 48). SME finds all legal mappings and then combines them to form all possible interpretations for the comparison. Selected interpretations correspond to the interpretation with the best relational structure.

If the base knowledge (or basedomain) and target knowledge share are lational structure, then significant inferences can be drawn from thebase domain in order to be transfered from base to target domain. This transfer is also controlled by structural constraints such as *Systematicity* principle. Gentner has argued that mapping ("analogyengine") is not influenced by knowledge. Therefore, SME simulates a mapping process that is independent of domain content, goals and context (Ripoll, 1993). However, this characteristic of SME is not compatible with what it is acknowledged about the influenced nature of human thought (Keane, Ledgeway, &Duff, 1994).

Analogical Constraint Mapping Engine

ACME uses parallel-constraint satisfaction method to construct a single, best interpretation of the comparison. This model is an interactive network. Three Constraints are implemented in ACME, namely *structural, similarity* and *pragmatic constraints*. In the network, a node represents a match between two predicates. For example,the match between SMART (steve) and ANGRY (fido) involves nodes representing the matches between SMART=ANGRY and steve=fido. Nodes are connected by excitatory and inhibitory links which implement the three constraints. The network runs until the activations of nodes settle into a stable state. The nodes in which activation exceeds a certain threshold arematches of the best interpretation. Mapping difficulty is measured by the number of cycles the network goes through before reaching the correct mapping.

This model has drawn our attention because it was one of the first model to take into account and examine pragmaticconstraint. Holyoak and Thagard (1989) contend that analogical mapping process could be influenced by pragmatic aspects of the base. According to these authors, "pragmatic" term concernselements which people assess to be important to reach a goal.

The aim of my thesis is to show that analogical mapping is strongly influenced by pragmatic information, namely thegoal.

Our assumption is that mapping process between target (t), knowledge (k) and base (b) knowledge progresses according to the goal *representation* that subjects want to reach.Unlike Holyoak and Thagard (1989), we assume that part of pragmatic information is played from target knowledge (t_k) and not from base knowledge (b_k):

$$M : t_k \rightarrow b_k$$

EXPERIMENT

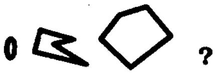
Our experiment was controled by computer.HyperCard 2.0© software had been used to carry out this experiment.

Materials

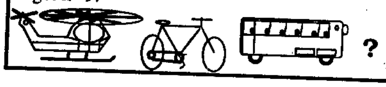
Ten target problems were composed of four termswhich the fourth one was missing. We have changed kind of terms in order to-make the experiment more attractive. We have displayed figures (or numbers), letters (or words), geometric shapes, and drawings terms.

Examples:

target n° 2: 3 9 27 ?
target n° 3: Bâton Belle Boeuf ?
target n° 6:



target n° 9:



Every target problem was matched with three base(a), (b), and (c) problems. Base problems were composed of four terms.

Only one relation was included in target problems where as two relations were included in base problems. Target relation waseither *belonging to a category* (eg., odd number) or *series of objects or events* (eg., increasing number).

Example:

Bases n° 2:	(a)	18	13	83	
	(b)	12,3	12,1	11,9	11,7
	(c)	10,2	11	21,2	32,2

The two relations of base problems were either *category* and *series* or *category* and *same surface* or *series* and *same surface*. "Surface" term means object properties shared by two situations and which are irrelevant to solve a problem: e.g., colors, shapes and so on. However, numerous empirical findings have shown that surface similarity facilitates retrieving process (see, e.g., Gentner & Landers, 1985; Holyoak & Kho,1987; Ripoll, 1998)

Procedure and task

Target problems were successively presented alone to the participants. Unlike classical paradigm, target problems were shown before base problems. The aim was to test our assumption according to which analogical mapping would be governed by goal representation of the target situation. Targets were displayed during two seconds and then were removed so that they should not be solved immediately. This time limitation allowed however subjects to encode terms of target problems.

Each target problem was once more presented with base problems (a), (b) and(c) in random order. With three target-base pair, participants were asked to assess whether the base was a support to solve the target problem. Participants clicked with cursor on *yes*or *no* button: it was the mapping task. Mapping times were recorded by the computer. After mapping target and bases, subjects gave answers to solve target problems. Verbal answers were typed and recorded.

	a	b	c	sum
yes	82	80	60	222
no	118	120	140	378

Chi-squared (2) = 6.349, $p < .0418$.

Table 1. Distribution of responses (yes/no), according tobases (a, b, c).

EXPECTATIONS

Mapping patterns

We expected participants to assess (a) and (b)bases to be more relevant than (c) bases. Accordingly, *yes*(y) responses would be linked with (a) and (b) bases and *no* (n) responses with (c):

a	b	c
yes	yes	no

Mapping times

We predicted that mapping times should take some time. Analogical mapping is a conscious process (Ripoll, 1992) that simultaneously developed between target situation and base situation.Therefore,this process has a high level time cost.

We expected participants to spend more time to conclude that base problemcould be a support to solve target problem than to conclude that base problem is not relevant. This prediction is associated with mapping pattern predictions.

Verbal answers

After the mapping task, subjects were asked to suggest answers to solve target problems. We predicted that verbal answers should be consistent with mapping patterns: if subjects click on *yes* button, then verbal answers should be matched with the fourth term of base problem.

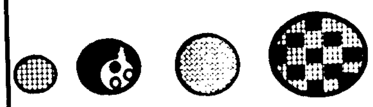
First observations

At this time, only 20 students of University of Provence took part voluntarily in the experiment.

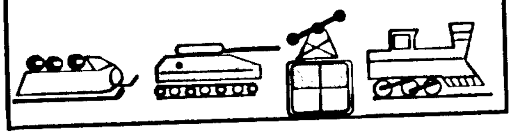
Mapping patterns

A descriptive analysis shows that subjects answer innegative form with bases (a) and (b).

drawings likebase (a) n°10 which was balls:



or like the cable-car. base (b) n°9:



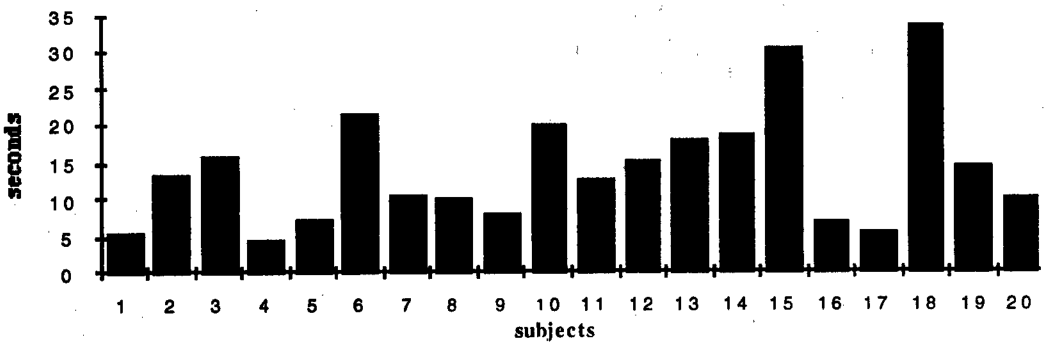


Figure 1. Mean mapping times taken by each subjects to assess if bases are, or are not a support to finish target problems.

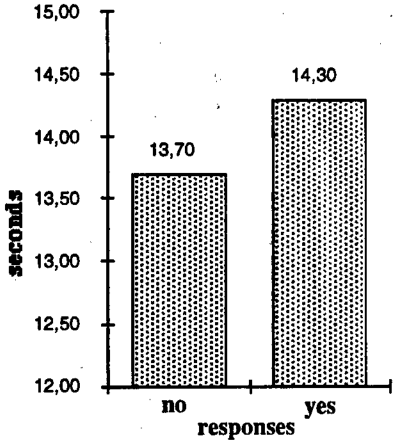


Figure 2. Means mapping times to clck on yes button (relevant base to solve target problems) or no button (irrelevant base).

This was observed whatever the kind of (figures, letters etc.) target-base pair.

These results are different from our expectations.Only base problems (c) are mainly refused as support to complete target problem.

There were eight possible patterns. Expected pattern(*yyn*) is not the most frequent: 14% whereas *nnn* pattern represents 29,5%.

yyy	yyn	yny	ynn	nyy	nyn	nny	nnn
17	28	17	28	3	32	16	59

Table 2. Distribution of pattern responses.

There is a lot of negative answers. Then, we have to think about the difficulty of the mapping task. In addition, a few participants said that they had difficulties to understand drawings likebase (a) n°10 which was balls:

But subjects' behaviour could be also implicated in this difficulty. We notice that subjects were looking for too complex relations whereas relations contained in materials were simple:increasing & decreasing; fast & slow; quadrilateral & ellipse, for example. Moreover, a few subjects expressed that they had removed out of their mind simplest relations because they thought that materials were designed with complex relations.

Mapping times

Mapping times show that mapping process takes a long time. The elapsed mean time was 13,9 seconds. In addition, there was alarge vari-

	a	**b**	**c**	
	yes	yes	no	target
coherent answers	15	18	0	0
surface similarity	2	2	0	0
other relations	0	3	0	5

Table 3. Distribution of verbal answers with regard the yyn pattern and respect to the categories of responses.

ability between subjects. For exemple, subject n°4 took 4,46 seconds to click on *yes* or *no* button and subject n°18 took 33,23 seconds (see figure 1).

As it was expected, subjects spent more time to assess that base problem could be a support to solve target problem than to conclude that base problem is not relevant (see figure 2).

However, the overall difference between *yes* and *no* responses is not significant. At present, we analyse more precisely mapping times of subjects whose pattern was: *yes yes* and *no* .

Verbal answers

Verbal answers given by participants were grouped together in three categories: *coherent answers*, answers based on *surface similarity*, and *otherrelation*. When we connect these categories with the eight patterns, we notice that,in general, subjects have proposed answers which were coherent with their patterns. When they thought a base was useful to solve a target problem,they gave an answer which was coherent with the fourth term of base problem. This result can alsobe observed with *yyn* pattern though the difficulty of the mapping task, and the time spent to match target and base problems.

Distribution of verbal answers with regard the yyn pattern and respect to the categories of responses

CONCLUSION

Intellectual honesty oblige us to be careful. First,the number of participants is insufficient and we have to change few drawings. Second, we have results which require more precise statistic analyses. However, first analyses of verbal answers would seem to indicate that analogical mapping process could be influenced by the target problems which were shown before base problems.

Another question concerns the lack of spontaneity of subjects. People are too centered on finding one solution. This experiment allows participants to befree in their answers. They are not instructed to be fast and they areasked to suggest as many responses as possible to solve target problems. It is important people feel free because,according to the answers proposed, we are in position to study how subjects perceive the goal of the situation where they are involved.

ACKNOWLEGMENTS

I would thank Christine Fiandino who has me heruseful english dictionary. Special thank to Nathalie Bonnardel (andMarie-Julie Bonnardel, she knows why) who has given me last and relevantadvices. I also thank Claude and Mireille Bastien.

REFERENCES

Bastien, C. (1997). *LesConnaissances de l'enfant à l'adulte*. Paris: Armand Colin.

Bastien-Toniazzo, M., Blaye, A., &Cayol, D. (1997). Résolution de probléme par analogie par des enfantsde grande section de maternelle. *L'Année Psychologique, 97*, 409-432.

Cauzinille-Marméche, E. (1990). Apprendre à utiliser desconnaissances pour la résolution d'un probléme: analogie ettransfert. *Bulletin de Psychologie*, Tome XLIV, 359.

Cauzinille-Marméche, E., Mathieu,J., & Weil-Barais, A. (1985). Raisonnement analogique et résolution deproblémes. *L'Année Psychologique, 85*, 49-72.

Falkenhainer, B., Forbus, K.D., &Gentner, D. (1986). Stucture-mapping engine. *Proceedings of theAnnual Conference of the American Association for ArtificialIntelligence*. Los Altos, CA: Mogan Kaufmann.

Falkenhainer, B., Forbus, K.D., &Gentner, D. (1989). Stucture-mapping engine. *ArtificialIntelligence, 41*, 1-63.

Gineste, M.D. (1997). *Analogie etCognition, Etude expérimentale et simulation informatique*.Paris: Presse Universitaire de France.

Gentner, D. (1983). Stucture-mapping: atheoretical framework for analogy. *Cognitive Science,7*, 155-170.

Gentner, D. (1988). Metaphor as Stucture-mapping: The Relational Shift.*Child Development, 59*, 47-159.

Gentner, D., & Landers, R. (1985).*Analogical reminding: A good match is hard to find.* Paper presented at the International Conference of Systems, Man & Cybernetics,Tuscon, AZ.

Goswami, U. (1992). *Analogical Reasoning in Children*, U.K., U.S.A.: Lawrence ErlbaumAssociates.

Goswami, U., & Brown, A.L. (1990).Higher-order structure and relational reasoning: contrasting analogical and thematic relations.*Cognition, 36*, 207- 226.

Holyoak, K.J., Junn, E.N., & Billman,D.O. (1984). Development of analogical problem-solving skill. *Child Development, 55*, 2042-2055.

Holyoak, K.J., & Kho, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*, 332-340.

Holyoak, K.J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*,295-355.

Keane, M., Ledgeway, T., & Duff, S. (1994). Constraint on AnalogicalMapping: a Comparison of Three Models. *Cognitive Science,18*, 387-438.

Piaget, J., Montangero, J., & Billeter, J. (1977). Les corrélats. InPiaget (Ed.),*L'Abstraction Réfléchissante*. Paris:Presses Universitaires de France.

Richard, J.F. (1990). *LesActivités Mentales: Comprendre, Raisonner,Trouver des Solutions*. Paris : Armand Colin.

Ripoll, T., 1993, *Rechercheen Mémoire d'un Probléme Analogue*. Thése de Doctorat.Université de Provence, Aix-Marseille I.

Ripoll, T. (1992). La recherche sur le raisonnement par analogie:objectifs, difficultés et solutions. *L'AnnéePsychologique, 92*, 263-288.

Ripoll, T. (1998). Why This Makes MeThinking of That. *Thinking and Reasoning, 4*, 15-43.

Savelli, J. (1993). *FacetteStatique et Dynamique de la Notion d'Analogie: Relation d'Analogie et Processus Analogiques*. Thése de Doctorat. Université de Droit et d'Economie et des Sciences, Aix-Marseille III.

THE NEUROANATOMY OF ANALOGICAL REASONING

**Charles M. Wharton[1], Jordan Grafman[2], Stephen K. Flitman[3],
Eric K. Hansen[4], Jason Brauner[5], Allison Marks[6], and Manabu Honda[7]**
[1]Language Section, 10/3C716, NIDCD, National Institutes of Health, Bethesda, Maryland USA
20892, cwharton@pop.nidcd.nih.gov; [2]Cognitive Neuroscience Section, 10/5C205, NINDS,
National Institutes of Health, Bethesda, Maryland USA 20892, jgr@box-j.nih.gov; [3]Cognitive
Neurology Section, Barrow Neurological Institute, 222 West Thomas Suite 304, Phoenix, AZ,
USA 85013, sflitman@mha.chw.edu; [4](same as [2]), eric.hansen@yale.edu; [5](same as [2])
jsbraune@midway.uchicago.edu; [6](same as [2]), amarks@helix.nih.gov; [7]Department of Brain
Pathophysiology, Kyoto University School of Medicine, 54 Shogoin, Sakyo, Kyoto 606-01 Japan,
mhonda@kuhp.kyoto-u.ac.jp

The distributed neural network that subserves analogical reasoning was identified using ^{15}O PET on 12 normal, high intelligence adults. Each trial presented during scanning consisted of a source picture of colored geometric shapes, a brief delay, and a target picture of colored geometric shapes. Analogous pictures did not share similar geometric shapes but did share the same system of abstract relations. Subjects judged whether each source-target pairing was analogous (analogy condition) or identical (literal condition). The results of the analogy-literal comparison showed left hemisphere activation in the inferior, middle, and medial frontal cortex, the inferior parietal cortex, and the superior occipital cortex. Based on converging evidence from neuropsychological and neuroimaging studies, we hypothesize that the inferior frontal and the inferior parietal cortices mediate analogical mapping.

THE NEUROANATOMY OF ANALOGICAL REASONING

Analogical mapping is important to understand because it is a cognitive ability necessary for explanation, learning, and categorization within virtually all forms of discourse. Although a considerable amount is known about its psychological aspects, extremely little is known about the neuroanatomical basis of analogical reasoning. To date, there have been no neuroimaging investigations of analogy with positron emission tomography (PET) or functional magnetic resonance imaging (fMRI), nor any focal lesion studies. However, a hypothesis about the neuroanatomical basis of analogical mapping can be made on the basis of neuropsychological studies of other forms of structure-driven reasoning (e.g., deduction) and ^{133}XE imaging experiments which have used analogical materials. On this basis we hypothesize that analogical mapping should be mediated by a distributed network based in the left prefrontal cortex and the left inferior parietal cortex. We report the results of a PET study that supports this hypothesis.

Reasoning and the brain

Because analogy theoretically shares many of the same representations and processes as logic and deduction (Halford, 1992), we can use neurological theories of deduction as a partial basis for neurological theories of analogical mapping. As reviewed in Wharton and Grafman (1998), an important distinction among cognitive theories of deduction is whether or not they focus on the influence of socially relevant content. *Content* refers to statements that imply a causation or social regulation (e.g., If one is to drink alcohol, one must be over eighteen). In contrast, a content independent statement implies relatively little relevant information (e.g., If there is an A on one side of a card, then there is a 4 the other side).

Clinical and neuroimaging studies appear to show that the left hemisphere conducts reasoning on the basis of formal logical operations whereas the right hemisphere and the medial ventral frontal cortex reason on the basis of experience. In Golding (1981), subjects were neurological patients with either no cerebral brain lesions, right hemisphere brain lesions, or left hemisphere brain lesions. These subjects were tested with a version of the Wason (1966) selection task. Subjects were shown cards that each had half of the top side masked. The unmasked side of each card showed either a circle, a diamond, a yellow patch, or a green patch. The task was to name the cards that would need to be unmasked to discover the truth of the rule, "whenever there is a circle on one half of the card there is yellow on the other half of the card." The rule would be falsified if the other side of the circle card showed green or if the other side of the green card showed a circle. Whereas only one left hemisphere lesioned patient and no control patients picked the circle and green cards, ten of the twenty right hemisphere lesioned patients surprising did better and picked these two cards. This finding points to the crucial role of the left hemisphere in deductive reasoning.

Additional evidence for the primary role of the left hemisphere in logic and deduction is provided by studies showing the difficulty that aphasics (especially with left posterior lesions) have in understanding even the simplest logic statements. Importantly, these studies indicate that right hemisphere lesioned subjects do not show general logical reasoning difficulties (Wharton & Grafman, 1998).

Ideally, in analogical reasoning, the objects and actions being mapped are much less significant than the structural relationships between these objects and actions (e.g., Gentner's (1983) "systematicity,"; Holyoak & Thagard's (1989) "isomorphism"). For example, in the Bohr planetary analogy of the atom, electrons are mapped to planets, not because of any physical or conceptual similarity, but because both revolve around a central body. Thus, it is likely that analogical reasoning, unless concerning topics with relevant content, is also dependent upon the left hemisphere.

Analogy and the brain

Although there has been little research into the neural basis of analogical reasoning, a number of studies have used analogical materials as a means of inducing verbal cognitive processing in subjects (Gur et al, 1994; Risberg, 1975). In these studies, a ^{133}Xe inhalation technique was used assess subjects' regional cerebral blood flow (rCBF) while they rested or solved four-term verbal analogies (e.g., kite is to air as raft is to a) fish, b) swimmer, c) duck, or d) water). These studies' hypotheses were not addressing analogical reasoning per se. Accordingly, designs were used that did not subtract out rCBF from cognitive activity not specific to analogical mapping (e.g., reading). Compared to a resting baseline, subjects solving analogy problems generally show more activation in the left than in the right hemisphere, particularly the posterior temporal and parietal cortices. Gur et al. (1994) noted that the analogical reasoning performance was significantly correlated with rCBF detected around the left inferior parietal cortex and so speculated that the left angular gyrus may be especially central to analogical reasoning. The left inferior parietal cortex has also been shown to be important to computational processes related to analogy such as arithmetic processing (Ardilla, 1993) and reasoning with spatial propositions (Hier et al., 1980). Thus, it is likely that the left inferior parietal cortex is an important part of the distributed neural network in the brain that mediates rule-based cognitive processes.

^{133}Xe studies using analogical materials have not shown significant activation in the left prefrontal cortex. However, various researchers have speculated that the dorsolateral prefrontal cortex (DLPFC) is specialized for mapping arguments to complex mental representations (Grafman, 1995; Holyoak & Kroger, 1995; Robin & Holyoak, 1995). Analogical mapping may be an emergent special case of this general property of the DLPFC. Also, several studies have reported that Broca's aphasics are impaired in logic and deduction (Wharton & Grafman, 1998) and a PET study of deduction reported that subjects solving deduction problems showed left prefrontal activation (Goel et al., 1997). Finally, given the amount of evidence in support of the view that regions in the left prefrontal cortex are responsible for syntactic language processing (Caplan, Hildebrandt, & Makris, 1996) and the fact that analogical mapping strongly resembles a syntactic process, it is likely that the left prefrontal cortex, as well as the left inferior parietal cortex, mediates analogy.

Method overview and hypothesis predictions

We used PET with ^{15}O labeled water to measure the rCBF of subjects performing an analogical match-to-sample task and a literal match-to-sample task. The literal task served as a comparison condition for the analogy task.

Visual objects were used as stimuli so that a large number of *novel* analogies could be created. (Although not explored as extensively as verbal analogical reasoning, visual analogical reasoning has been studied both with behavioral experiments (Gick, 1985; Goswami, Brown, Mulholland, Pellegrino, & Glaser, 1980) and with computational modeling (Goldstone, 1994; Thagard, Gochfeld, & Hardy, 1992)). Stimuli consisted of groups of three to five colored, geometric shapes such as circles and stars that were framed by a larger geometric shape such as a square, circle, rectangle, diamond, or triangle (see Figures 1 and 2). All objects within these frames could be easily labeled verbally (e.g., "rectangle").

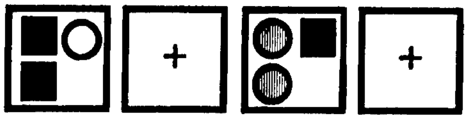


Figure 1. Example of stimuli for a correct trial sequence in the analogy condition.

As shown in Figure 1, individual trials consisted of the sequential presentation of a source picture (3 s display), a fixation cross (intratrial delay), a target picture (3 s display), and then another fixation cross (intertrial delay). In the *analogy* conditions, subjects indicated whether the target picture was an analog of the source picture. In each correct trial, the source and target pictures contained different objects but shared the same system of relations. In each incorrect trial, one object in the target was mismatched to its corresponding object in terms of its spatial relationship (i.e., position) or object relationship (i.e., shape, texture, or color) (see upper right two panels of Figure 2). In the literal conditions, subjects indicated whether the target picture was an exact match of the source picture. In each correct trial, the source and target pictures were identical, whereas in each incorrect trial, one object in the target was a different object or was spatially displaced (see bottom right two panels of Figure 2).

We used a 2 (Similarity: analogical, literal) x 2 (Intratrial Delay: immediate, delay) design that produced four conditions. In the delay analogy and delay literal conditions, the intratrial and intertrial delays were 3000 ms and 500 ms, respectively. In the immediate analogy and immediate literal conditions, the intratrial and intertrial delays were 100 ms and 3400 ms, respectively. The delay and immediate conditions were designed so that when compared to each other, rCBF activation would be shown specific to holding the mental representations

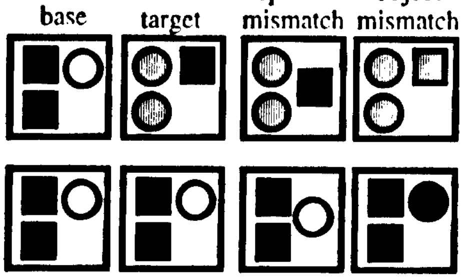


Figure 2. Correct and incorrect trials for the analogy condition (top row) and the literal condition (bottom row).

of the source pictures in working memory. The analogy and literal conditions were designed so that when compared to each other, rCBF activation would be shown for brain regions engaged in analogical mapping. Given that our materials require subjects to perceive spatial-object analogies, it is relevant to note Heir et al.'s (1980) examination of three semantic aphasics, two with infarctions of the left parieto-occipital junction and one with a bilateral hemorrhage of the parieto-temporo-occipital junction. Whereas these patients could use abstract words such as *crystallized, saccharin, immature,* and *decisive*, they could not correctly follow commands using spatial prepositions such as *beside, under, behind, before,* or *away from,* nor comprehend simple logico-grammatical relationships. Hier et al. concluded that the left temporo-parieto-occipital region subserves perception of spatial relationships (see Farah, 1995, D'Esposito et al. , 1997). Thus, we predicted that the analogy-literal comparison would reveal activation in the left inferior parietal cortex, adjacent areas in the left occipital cortex, and the left prefrontal cortex.

METHOD

Subjects

Subjects were 6 females and 6 males, all right-handed (mean age and years of education, 26 years and 18 years, respectively). Subjects' mean scaled scores on both the WAIS (Weschler, 1991) vocabulary and block design subscales were above average (13 and 12, respectively).

Materials and Apparatus

All source pictures appeared, across subjects, in analog and literal conditions (see left column of Figure 2).

For our stimuli, *spatial relations* refers to categorical predicates describing the relative spatial positions of all objects in a picture (e.g., *diagonal_to* (blue (oval1), blue (oval3)). *Object relations* refers to categorical predicates describing the relative shape, color, size, and texture of objects to each other (e.g., *three_of_a_kind* (blue (oval1), brown (oval2), blue (oval3)). The *system of object and spatial relations* refers to the combinations of predicates required to fully describe each picture (e.g., *three_of_a_kind* (*diagonal_to* (blue (oval1), blue (oval3)), (*above* (brown (oval2), blue (oval3)), etc.). We assume that object and spatial predicates can be represented in verbal, visual, or both modalities.

The following factors influenced the design of stimuli for incorrect trials:

1. We wanted subjects to map each picture's system of object and spatial relations. For 50% of incorrect analogy trials, one object in the target picture was spatially mismatched to an object in the source picture that it correctly matched for color, shape, and size relations (see middle oval in the upper middle right panel of Figure 2). In the other incorrect trials, one object in the target picture was mismatched in terms of object relations to one object in the source picture that it spatially matched (see triangle in the upper far right panel of Figure 2).

2. Literal trials were designed to subtract activation from the analogy trials in statistical analysis. Accordingly, except for analogical reasoning, we wanted to minimize the differences in the cognitive processes that were used in performance of analogy and literal trials. Incorrect *literal* trials were similar to incorrect analogy trials in that one object in the target picture was either mismatched in terms of its previous spatial position or object characteristics (see lower right two panels of Figure 2). A side effect of this manipulation is that incorrect literal trials were likely easier to detect than incorrect analogy trials. An alternative way of constructing incorrect literal trials would have been to make them equivalent in difficulty to incorrect analogy trials by making relatively subtle object and spatial changes between incorrect literal base and target images. However, such a materials manipulation would possibly require sub-

jects to use qualitatively different encoding and comparison processes in the literal condition as compared to what they would use in the analogy condition.

PET scans were performed using a Scanditronix PC2048-15B [Uppsala, Sweden], which collected 15 contiguous planes with 2 mm x 2 mm x 6.5 voxels resolution.

Procedure

After 40 min of pretraining, each subject was scanned twice in each condition. All presented pictures were seen only once, and an equal number of false and true trials were presented in each scan. Presentation order of the four conditions was counterbalanced across subjects, and all source-target pairings were seen equally in delay and immediate conditions. To control for neural activation from eye movement, each picture was displayed separately to subjects (see Fig. 1).

Each subject's head was secured with a conforming plastic mask and positioned for scans from 14 mm to 111.5 mm above the canthomeatal line. A transmission scan was obtained with a rotating ^{68}Ge/^{68}Ga source. Each scan resulted from an intravenous bolus of 37 mCi $H_2{}^{15}O$, for a 60 sec period beginning 13-16 s after bolus.

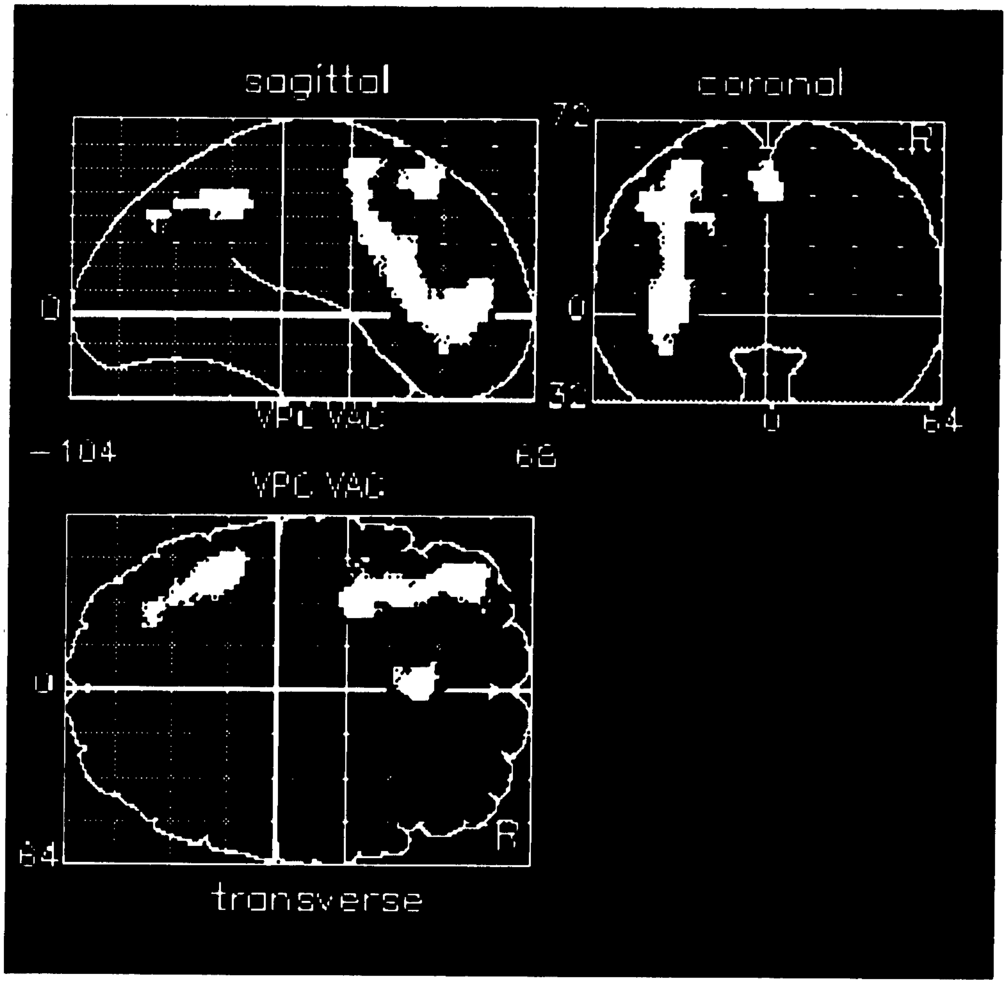


Figure 3. Brain regions activated in the analogy-literal comparison.

RESULTS

Behavioral measures

Mean differences were tested with a two-way within-subjects ANOVA. As compared to their performance during scanning in the literal condition, subjects' performance during scanning in the analogy condition was slower (1415 vs. 984 ms.; $F(1, 11) = 128.49$, $p <.0001$) and less accurate (respectively, 97% vs. 87%, $F(1, 11) = 127.08$, $p <.0001$). Subjects' accuracy rates ranged between .81 and .94 in the analogy condition and between .91 and 1.00 in the literal condition. For accuracy rates, the main effect of delay and the interaction of similarity by delay were not significant (both $F < 1$).

Functional measures

Scans were realigned to correct for head movement, then normalized to the Talairach and Tournoux anatomic space (Talairach & Tournoux, 1988). Smoothing was done with a 20 mm x 20 mm x 12 mm Gaussian filter to reduce mismatch due to anatomic variation. Subject-specific ANCOVA was used to discount variations in overall intensity between scans. Within-group comparisons of rCBF were produced by statistical parametric mapping (SPM95; Wellcome Department of Cognitive Neurology, London, UK; Frackowiak, & Friston, 1994) with tests of significance for the size of the activated region (Friston, Worsley, Frackowiak, Mazziotta, Evans, 1993-1994). Regions of interest (ROIs) were defined by a threshold of $Z=3.09$ for each contrast between conditions. For each ROI, statistical probabilities obtained included a p value ($a = .05$) for whether the ROI's peak intensity difference was significant and also for the ROI's spatial extent representing the probability that the clustered voxels comprising the ROI arose by chance.

Figure 3 displays a three axis SPM plot of the analogy-literal comparison (left hemisphere is left on transverse and coronal views; frontal areas are to right on sagittal and transverse views). As shown in Figure 3, the analogy-literal comparison indicated significant rCBF activation in the medial frontal cortex and in left hemisphere regions including the DLPFC and a parietal-occipital area. Specific locations of local maxima are shown in Table 1. The DLPFC region had a local maxima in the middle frontal gyrus (BA 6) as well as other significant maxima in the inferior frontal gyrus (BA 10, 44, 45, 46). The medial frontal cortex region had a local maxima in the superior frontal gyrus (BA 8). The parietal-occipital region had a local maxima in the inferior parietal lobule (BA 40) as well as other significant maxima in the inferior parietal lobule (BA 7, 40), and the superior occipital region (BA 19).

Neither the main effect of delay, nor the interaction of delay and analogy revealed significant activation.

DISCUSSION

The results of the PET scans indicate that relative to when performing literal matching, subjects performing analogical matching utilize a network consisting of the left inferior and middle prefrontal cortices, the medial frontal cortex, the left inferior parietal lobule, and the left superior occipital cortex. These results are noteworthy because they are the first to have come from an imaging study specifically designed to localize analogical reasoning. Additionally, our results add converging evidence to the idea that content-independent reasoning is mediated by the left hemisphere (Wharton & Grafman, 1998). Finally, these results support the theorized role of the frontal cortex in reasoning (Grafman, 1995; Holyoak & Kroger, 1995).

There are two alternative explanations for our results. First, subjects may have been looking only for simple spatial "popout" in the analogy condition. However, if one judged that a match had occurred unless one detected a spatial mismatch, the maximum obtained correct rate would be (1*.25 [spatial mismatches] + 0*.25 [object mismatches] + 1*.50 [correct matches]) = .75. The lowest accuracy rate of any subject in the analogy condition was .81. Further, given the fact that base pictures were complex and novel, as well as perceptually different from their targets, looking for popout with both object and spatial relations would require full analogical mapping anyway. Second, because subjects' activation was almost solely in the left cerebral cortex, the cause of this activation may have been due entirely to phonological working memory (Baddeley, 1992). However, experiments have demonstrated that subjects' performance in verbal deduction problems is not significantly affected by verbal rehearsal (Hitch & Baddeley, 1976; Gilhooly, Logie, Wetherick, & Wynn, 1993; Toms, Morris, & Ward, 1993).

A consequence of constructing incorrect literal trials similar to analogy trials is that subjects were more accurate in the literal condition than in the analogy condition. Although we believe that the significant activation differences of the analogy-literal comparison are the result of analogical processing, some of the activation differences may also reflect the additional attention needed to perform in the analogy condition.

Besides activation in the left anterior and posterior regions, the analogy-literal comparisons also revealed activation in the dorsal medial frontal cortex. Research with monkeys has shown that this area is involved with spatial attention processes (Lee & Tehovnik, 1995). Thus, dorsal medial frontal activation may have been due to extra spatial processing required in the analogy condition, spatial and object analogical mapping, or both.

The working memory comparison (i.e., delay - immediate) may not have shown significant activation because the process of holding stimuli in working memory for 3 s was not inherently a demanding enough task to produce significant activation. Alternatively given subjects' extensive pretraining, subjects may have become so practiced at keeping mental representations of the stimuli in mind that associated brain activations fell below detectable levels.

CONCLUSION

Our results support the hypothesis that the left prefrontal inferior parietal cortices are especially central to analogical mapping. Our findings are especially important because they are the first to localize the crucial cognitive processes required for analogical mapping to specific brain regions as well as demonstrating that analogical mapping is a tractable topic for neuroimaging investigation. Our results should provide encouragement for more focused neuroanatomical studies of analogical reasoning.

REFERENCES

Ardilla, A. (1993). On the origins of calculation abilities. Behavioral Neurology, 6, 89-97.

Baddeley, A. (1992). Working memory. Science, 255, 556-559.

Barnden, J. A. (1994). On the connectionist implementation of analogy and working memory matching. In J. A. Barnden & K. J. Holyoak (Eds.), Advances in connectionist and neural computation theory, Vol. 3: Analogy, metaphor, and reminding, 327-374. Norwood, NJ: Ablex.

Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. Brain, 119, 933-949.

D'esposito, M., Detre, J. A., Aguiree, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J., & Farrah, M. (1997). A functional MRI study of mental image generation. Neuropsychologia, 35, 725-730.

Farah, M. J. (1995). The neural bases of mental imagery. In M. S. Gazzaniga (Ed.), The cognitive neurosciences (pp. 963-975). Cambridge, MA: MIT.

Frackowiak, R. S., & Friston, K. J. (1994). Functional neuroanatomy of the human brain: positron emission tomography—a new neuroanatomical technique. Journal of Anatomy, 184, 211-225.

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., Evans, A. C. (1993-1994). Assessing the significance of focal activations using their spatial extent. Human Brain Mapping, 1, 210-220.

Gentner, D. (1983). Structure-mapping: A theoretical framework. Cognitive Science, 7, 155-170.

Gick, M. L. (1985). The effect of a diagram retrieval cue on spontaneous analogical transfer. Canadian Journal of Psychology, 39, 460-466.

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. Memory & Cognition, 21, 115-124.

Goel, V., Gold, B., Kapur, S., & Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. Neuroreport, 8, 1305-1310.

Golding, E. (1981). The effect of unilateral brain lesion on reasoning. Cortex, 17, 31-40.

Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 3-28.

Goswami, U., & Brown, A. L. (1990). Higher-order structure and relational reasoning: Contrasting analogical and thematic relations. Cognition, 36, 207-226.

Grafman, J. (1995). Similarities and distinctions among current models of prefrontal cortical functions. In J. Grafman, K. J. Holyoak, & F. Boller (Eds.). Structure and functions of the human prefrontal cortex. Annals of the New York Academy of Sciences, 769, 337-368.

Gur, R. C., Ragland, J. D., Resnick, S. M., Skolnick, B. E., Jaggi, J., Muenz, L., & Gur, R. E. (1994). Lateralized increases in cerebral blood flow during performance of verbal spatial tasks: Relationship with performance level. Brain & Cognition, 24, 244.

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. Human Development, 35, 193-217.

Hier, D. B., Mogil, S. I., Rubin, N. P., & Komros, G. R. (1980). Semantic aphasia: a neglected entity. Brain and Language, 10, 120-131.

Hitch, G. J., & Baddeley, A. D, (1976). Verbal reasoning and working memory. Quarterly Journal of Experimental Psychology, 28, 603-621.

Holyoak, K. J., & Kroger, J. K. (1995). Forms of reasoning: Insight into prefrontal functions? In J. Grafman, K. J. Holyoak, & F. Boller (Eds.), Structure and functions of the human prefrontal cortex Annals of the New York Academy of Sciences, 769, 253-263.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. Cognitive Science, 13, 295-355.

Lee, K., & Tehovnik, E. J. (1995). Topographic distribution of fixation-related units in the dorsomedial frontal cortex of the rhesus monkey. European Journal of Neuroscience, 7, 1005-1011.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. Cognitive Psychology, 12, 252-284.

Risberg, J., Halsey, J. H., Wills, E. L., & Wilson, E. M. (1975). Hemispheric specialization in normal man studied by bilateral measurements of the regional blood flow. Brain, 98, 511-524.

Robin, N., & Holyoak, K. J. (1995). Relational complexity and the functions of prefrontal cortex. In M. S. Gazzaniga (Ed.), The cognitive neurosciences (pp. 987-997). Cambridge, MA: MIT.

Talairach, J., & Tournoux, P. (1988). Co-Planar Stereotaxic Atlas of the Human Brain. New York: Thierne.

Thagard, P., Gochfeld, D., & Hardy, S. (1992). Visual analogical mapping. Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society (pp. 522-527). Hillsdale, NJ: Erlbaum.

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. Quarterly Journal of Experimental Psychology, A, 46, 679-699.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), New horizons in psychology (Vol. 1). Harmondsworth, UK: Penguin.

Wechsler, D. (1991). Wechsler adult intelligence scale - III. New York: Psychological Corporation.

Wharton, C. M., & Grafman, J. G. (1998). Reasoning and the brain. Trends in Cognitive Science, 2, 54-59.

WHY MONKEYS AND PIGEONS, UNLIKE CERTAIN APES, CANNOT REASON ANALOGICALLY

Roger K. R. Thompson
Franklin & Marshall College
Lancaster, PA. U.S.A.

David L. Oden
La Salle University
Philadelphia PA U.S.A.

ABSTRACT

Language-training, or prior experience with arbitrary symbols for the abstract concepts "same and different", appears to be necessary before chimpanzee or child can judge different pairs of objects or patterns to be analogically the same. Comparable training with symbols for "same and different", however, does not enable macaque monkeys to judge the analogical equivalence of stimulus pairs. Why should this be? There is, after all, good evidence that monkeys and pigeons can judge whether objects or events are the same on the basis of physical identity or membership in a common class or category. Unlike the chimpanzees and children, however, neither adult nor infant macaque monkeys spontaneously perceive the analogical identity of relations-between-relations. These results support the hypothesis that representational re-coding of abstract relations via symbols enable child and chimpanzee to explicitly express that which they, if not monkeys, perceive implicitly early in life.

Analogical Judgments of Similarity are a hallmark of human reasoning and intelligence (Spearman, 1923; Sternberg, 1977). Similarity judgments can be based solely on physical identity or the degree of resemblance between categorical attributes. Analogies, however, entail judgments about the equivalence of higher-order relational structures and representations that need not physically resemble one another (Gentner & Markman, 1997; Goswami, 1991; Holyoak & Thagard, 1997).

Recent research indicates that early in life humans and chimpanzees have perceptual and cognitive precursors for the development of higher level analogical information processing abilities that are not shared by adult or infant macaque monkeys (Thompson, 1995; Thompson & Oden, 1996). Furthermore, some form of re-coding via language or analogous symbolic systems catalyses the explicit expression of these implicit competencies in problem solving tasks involving analogical reasoning by both natural and artificial learning systems (Thompson, Oden, & Boysen, 1997; Clark & Thornton, 1997).

IMPLICIT AND EXPLICIT KNOWLEDGE ABOUT ANALOGICAL RELATIONS IN CHIMPANZEES AND CHILDREN.

Language-naive chimpanzees and pre-linguistic human infants perceive relations (identity or nonidentity) to be the same or different as measured by either visual gaze or object handling in preference-for-novelty tasks like 'paired-comparison' and 'habituation/dishabituation'. However, both non- or pre-linguistic species fail to explicitly judge the analogical equivalence of one identity relation (AA) with another identity relation (BB), and one nonidentity relation (CD) with another (EF) (Oden, Thompson, & Premack, 1990; Tyrrell, Stauffer & Snowman, 1991; Tyrrell, Zingaro, & Minard, 1993). Note that in this example, and for the remainder of this paper letters (e.g., AA & CD) are used only for expository purposes in lieu of the actual or digitized stimulus objects employed.

Only those humans and chimpanzees exposed to a regime of language or symbolic token training can judge abstract relations-between- relations as being the same or different (House, Brown & Scott, 1974; Premack 1978; 1983a, 1983b; Thompson, Oden & Boysen, 1997). For example, this capacity is revealed in conceptual matching-to-sample tasks. In this problem a chimpanzee or child is correct if they match a pair of shoes with a pair of apples, rather than to a paired eraser and padlock. Likewise, they are correct if they match the latter nonidentical pair with a paired cup and paperweight. The conceptual matching-to-sample task can be conceived of as a nonlinguistic analogy problem involving a single abstract relationship of same or different. Gillan, Premack & Woodruff (1981) demonstrated that a language trained chimpanzee - Sarah - who matched conceptually also succeeded in completing partially constructed analogies involving complex geometric forms and functional relationships. More recently, Oden, Thompson & Premack (in preparation) further demonstrated that this same chimpanzee could not only complete, but also construct, analogies spontaneously from a randomized grouping of geometric elements.

These findings imply that language or symbol training does not instill propositional knowledge about abstract relations of the type described above, but it does appear necessary for the explicit expression of such knowledge in equivalence judgment tasks. The implication then is that experience with external symbol structures and experience using them transforms the shape of the computational spaces that must be negotiated in order to solve certain kinds of abstract problems. This finding dovetails with the independent demonstration by Clark and Thornton (1997) that standard connectionist learning by artificial intelligent systems runs aground in exactly the same class of tasks used with the child and chimpanzee, unless the net is provided with some external means of reducing the search space.

MONKEYS DEMONSTRATE NEITHER IMPLICIT NOR EXPLICIT KNOWLEDGE ABOUT ANALOGICAL RELATIONS.

The provision of such 'external means' via symbol training with tokens does not enable macaque monkeys to judge the analogical equivalence of stimulus pairs (Washburn, Thompson & Oden, 1997; ms. in preparation). "Symbol" sophisticated monkeys were trained to choose "Circle" following an identity pair (AA—O) and to choose "Triangle" following a nonidentity pair (CD—/_\). Then they generalized this ability to novel identity (BB) and nonidentity (EF) stimulus pairs. Nevertheless, as shown in figure 1, unlike chimpanzees with the same experience (Thompson, Oden & Boysen, 1997), the monkeys still failed to match AA with BB and CD with EF above chance levels despite their success on physical matching problems. Why should this be? Thompson & Oden (1996) demonstrated that contrary to ape and child, adult macaque monkeys are perceptually insensitive to analogical equivalencies of a propositional nature. Hence, the circle and triangle tokens could not acquire symbolic meaning as was the case for chimpanzees. Instead the circle and triangle token were restricted to functioning simply as choice alternatives signaled by the preceding physical equivalence judgment that 'A is A' or 'C is not D'

Adult rhesus macaque monkeys do not spontaneously perceive analogical or relational identity when tested using the same preference for novelty procedures employed with the chimpanzees and human infants (Thompson, Oden, & Gunderson, 1997). Thus far, this disparity holds true regardless of the task (paired-comparison & habituation/dishabituation) and hence time available for information processing, or whether visual gaze or object handling is the dependent measure (Chaudhri, Ghazi, Thompson & Oden, 1997; Thompson, 1995; Thompson & Oden, 1996; Thompson, Oden, Boyer, Coleman, & Hill, 1997). Nevertheless, regardless of the dependent measure, the same animals give every indication that they perceive objects to be the same or different based on physical properties alone.

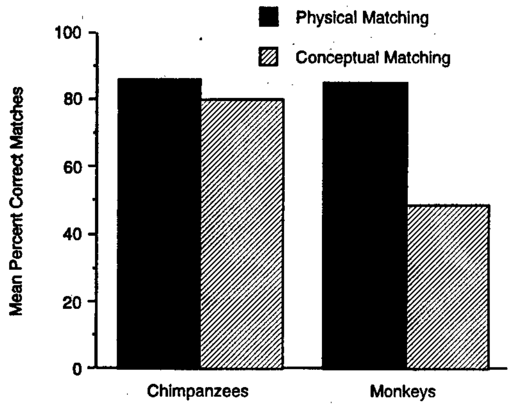


Figure 1. Percent correct performances for physical (i.e. object) and conceptual (i.e., analogical relations-between-relations) in matching-to-sample (MTS) tasks by chimpanzees and macaque monkeys previously trained with symbols for "same" and 'different". Data for chimpanzees derived from Thompson, Oden & Boysen (1997). Data for monkeys derived from Washburn, Oden, & Thompson (1997).

Recent data collected from infant macaques further indicate that these results are not simply a function of age (Maninger, Gunderson, & Thompson, 1997). As shown in figure 2, 7-week-old pigtailed macaque infants, like the adult macaques, but in contrast to their human counterparts, fail to recognize abstract relations on a visual paired-comparison measure. This is the first study using the familiarity-novelty paradigm in Gunderson's laboratory that has shown a discontinuity in perceptual-cognitive development between macaque and human infants (Grant-Webster, Gunderson & Burbacher, 1990; Gunderson, Rose & Grant-Webster, 1990; Sackett, Gunderson & Baldwin, 1982).

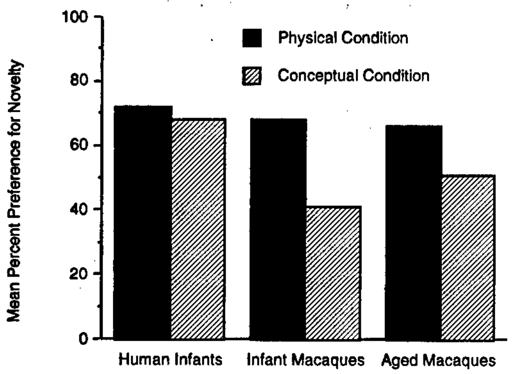


Figure 2. Percent preferences for physical/object and conceptual/relational novelty in visual paired-comparison tasks. Data for human infants derived from Tyrrell et al., (1991). Data for macaque monkey infants and adults derived from, respectively, Maninger, Gunderson, & Thompson (1997), and Thompson, Oden, & Gunderson, (1997).

CONCLUSIONS

Taken together all the above findings imply that analogical reasoning in natural, and possibly artificial, agents cannot emerge from a **tabula rasa**. Rather, as suggested also by Clark and Thornton's work (1997), the facilitative effects of language and symbol training on analogical reasoning can only operate upon pre-existing perceptual competencies. This restructuring of input/output spaces permits the establishment of new similarity or neighborhood relations between stimuli.

ACKNOWLEDGEMENTS

Preparation of this paper was supported by Funds from the National Science Foundation (NSF-IBN 9420756).

REFERENCES

Chaudhri, N., Ghazi, L., Thompson, R. K. R., & Oden, D. L. (1997). *Do monkeys perceive abstract relations in handled object pairs?* Paper presented at the annual meeting of the Eastern Psychological Association.

Clark, A., Thornton, C. (1997). Trading Spaces: Computation, representation, and the limits of uniformed learning. *Behavioral and Brain Sciences, 20,* 57-90.

Gentner, D. & Markman, A. B. (1997). Structural mapping in analogy and similarity. *American Psychologist, 52,* 45-56.

Gillan, D. D., Premack, D., & Woodruff, G. (1981). Reasoning in the chimpanzee: I. Analogical reasoning. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 1-17.

Goswami, U. (1991). Analogical reasoning: What develops? A review of research and theory. *Child Development, 62,* 1-22.

Grant-Webster, K. S., Gunderson, V. M., & Burbacher, T. M. (1990). Behavioral assessment of young nonhuman primates: Perceptual cognitive development. Neurotoxicology *& and Teratology, 12,* 543-546.

Gunderson, V. M., Rose, S. A., & Grant-Webster, K. (1990). Cross-modal transfer in high-and low risk infant pigtailed monkeys, *Developmental Psychology, 26,* 576-581.

Holyoak, K.J. & Thagard, P. (1997). The analogical mind. *American Psychologist, 52 (1),* 35-44.

House, B. J., Brown, A. L., & Scott, M. S. (1974). Children's discrimination learning based on identity or difference. In H. W. Reese (Ed.), *Advances in child development and behavior, Vol. 9,* New York: Academic.

Maninger, N., Gunderson, V. M., & Thompson, R. K. R. (1997). Perception of identity and difference relations in infant pigtailed macaques (*macaca nemestrina*). Paper to be presented at the Annual Meeting of the American Primatogolical Society in June.

Oden, D. L., Thompson, R. K. R., & Premack, D. (1990). Infant chimpanzees (*Pan troglodytes)* spontaneously perceive both concrete and abstract same/different relations. *Child Development, 61,* 621-631.

Premack, D. (1978). On the abstractness of human concepts: Why it would be difficult to talk to a pigeon. In S. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior.* Hillsdale, NJ: Erlbaum Associates.

Premack, D. (1983a), Animal cognition. *Annual Review of Psychology, 34,* 351-362.

Premack, D. (1983b). The codes of man and beast. *The Behavioral and Brain Sciences, 6,* 125-137.

Sackett, G., Gunderson, V. M., & Baldwin, D, (1982). Studying the ontogeny of primate behavior. In J.L., Fobes and J. E. King (Eds.) *Primate behavior* (pp. 135-169). N.Y.: Academic Press.

Spearman, C. (1923). *The nature of "intelligence" and the principles of cognition.* London: Macmillan.

Sternberg, R.J. (1977). *Intelligence, information processing, and analogical reasoning.* Hillsdale, NJ: Erlbaum.

Thompson, R. K. R. (1995). Natural and relational concepts in animals. In: H. Roitblat & J. A. Meyer (Editors), *Comparative Approaches to Cognitive Science* (pp.175-224). Cambridge, MA: Bradford / MIT Press.

Thompson, R. K. R., Oden, D. L. & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal behavior processes, 23,* 31-43.

Thompson, R. K. R., Oden, D. L., Boyer, B., Coleman, J. F., & Hill, C. C. (1997). *Test for the perception of abstract relational identity and nonidentity by macaque monkeys in an habituation/dishabituation task.* Paper presented at the annual meeting of the Eastern Psychological Association.

Thompson, R. K. R., Oden, D. L., & Gunderson, V. M. (1997). *Adult and infant monkeys do not perceive abstract relational similarity.* Paper presented at the 38th Annual Meeting of the Psychonomic Society. Philadelphia, PA.

Thompson, R. K. R., & Oden, D. L. (1996). A profound disparity revisited: Perception and judgment of abstract identity relations by chimpanzees, human infants, and monkeys. *Behavioral Processes, 35,* 149-161.

Thompson, R. K. R., & Oden, D. L. (1993). "Language training" and its role in the expression of tacit Tyrrell, D. J., Stauffer, L. B., & Snowman, L. G. (1991). Perception of abstract identity/difference relationships by infants. *In-*

fant behavior and Development, 14, 125-129.

Tyrrell, D. J., Zingaro, M. C., & Minard, K. L. (1993). Learning & transfer of identity-difference relationships by infants. *Infant Behavior and Development, 16,* 43-52.

Washburn, D. A., Thompson, R. K. R. & Oden D. L. (1997). Monkeys trained with same/different symbols do not match relations. Paper presented at the 38th Annual Meeting of the Psychonomic Society. Philadelphia, PA.

The Effect of Language on Similarity: The Use of Relational Labels Improves Young Children's Performance in a Mapping Task

Mary Jo Rattermann
Department of Psychology Whitely Psychology Labs Franklin and Marshall College
Lancaster, PA 17604 M_Rattermann@ACAD.FANDM.EDU

Dedre Gentner
Department of Psychology Northwestern University 2029 Sheridan Ave.
Evanston, IL 60208 gentner@nwu.edu

INTRODUCTION

The ability to use relational similarity is considered a hallmark of sophisticated thinking; it plays a role in theories of categorization, inference, transfer of learning and generalization (Gentner & Markman, 1997; Halford, 1993; Holyoak & Thagard, 1995; Novick, 1988; Ross, 1989). However, young children often fail to notice or use relational similarity (Gentner, 1988; Gentner & Rattermann, 1991; Goswami, 1993; Halford, 1993). For example, when given the metaphor "plant stems are like drinking straws" 5-year-old children focus on the common object similarities, commenting that "They are both long and thin," whereas 9-year-olds focus on the relational commonality that "They both carry water" (Gentner, 1988).

This relational shift in children's use of similarity—a shift from early attention to common object properties to later attention to common relational structure—has been noted across many different tasks and domains (Gentner & Rattermann, 1991; Halford, 1993). For instance, Gentner and Toupin (1986) presented children with a story mapping task in which object similarity and relational similarity were *cross-mapped*: that is, similar objects were placed in different relational roles in the two scenarios, so that the plot-preserving relational correspondences were incompatible with obvious object-based correspondences. Under these conflict conditions, 6-year-old children were unable to preserve the plot structure in their mapping, although they could transfer the story plot accurately when given similar characters in similar roles. Older children (9-years-old) could maintain a focus on the relational structure and transfer the plot accurately despite competing object matches. There is evidence that this shift from objects to relations is based on gains in knowledge (Brown, 1989; Goswami, 1993; Kotovsky & Gentner, 1996; Rattermann & Gentner, in press), although maturational changes may also play a role (Halford, Wilson, Guo, Gayler, Wiles & Stewart, 1995).

Children's ability to carry out purely relational comparisons improves markedly across development. Yet even very young children can reason analogically under some circumstances (Crisafi & Brown, 1986; Kotovsky and Gentner, 1996). For example, Gentner (1977) demonstrated that preschool children can perform a spatial analogy between the familiar base domain of the human body and simple pictured objects, such as trees and mountains. When asked, "If the tree had a knee, where would it be?," even 4-year-olds (as well as 6- and 8-year-olds) were as accurate as adults in performing the mapping of the human body to a pictured object, even when the orientation of the tree was changed or when confusing surface attributes were added to the pictures.

What factors impede or promote the perception of common relational structure? According to structure-mapping theory (Gentner, 1983, 1989; Gentner & Markman, 1997) an analogy is the mapping of knowledge from one domain (the base) to another domain (the target) in which the system of relations that holds among the base objects also holds among the target objects. When adults interpret an analogy, the correspondences between base and target objects are based on common roles in the matching relational structures; the corresponding objects in the base and target do not have to resemble each other. However, although the final interpretation of an analogy is determined by relational similarity rather than by object similarity, we hypothesize that in the actual process of computing an analogy both object similarity and relational similarity are at work (Falkenhainer, Forbus, & Gentner, 1990; Halford, Wilson, Guo, Gayler, Wiles & Stewart, 1995; Holyoak & Thagard, 1989; Hummel & Holyoak, 1997; Keane & Brayshaw, 1988).

A natural consequence of the structure-mapping view is that knowledge of relations plays a crucial role in the mapping process; if the child (or adult) has not represented the relations that hold within the domain then the matches formed will be based upon common object similarity rather than common relational similarity. Thus as domain knowledge increases, so does the likelihood that the child's comparisons will be based on common relational structure.

In summary, the ability to use relational similarity is sensitive to changes in the chil d's knowledge base. With increasing knowledge of the relationships in a domain, children become more able to understand and produce purely relational matches. This brings us to the issue of interactions between language and thought.

Language and Relational Similarity

We propose that language may interact with the development of analogical ability by serving as an invitation to seek likeness—to make comparisons. The word-learning studies of Markman, Waxman, and others have shown that when children are taught a new object term they assume very that the word applies to things of like kind (Imai, Gentner & Uchida, 1994; Markman, 1989; Waxman & Gelman, 1986). However, this work has focused on noun learning. We propose that the acquisition of relational language promotes the development of analogy by inviting children to notice and represent higher-order relational structure (Gentner & Medina, 1997). So far, the evidence on this issue is rather scant, although Kotovsky and Gentner (1996) found that 4-year-olds were better able to perceive cross-dimensional perceptual matches—e.g., symmetry of size compared to symmetry of shading—when they had previously been taught a relational label—"even"—to identify the relation of symmetry.

The Present Studies

In these experiments we tested whether children's relational performance can be improved by the introduction of relational labels. The basic task used in these experiments was a *cross-mapping* search task in which object similarity and relational similarity were in conflict so that a response based on one type of similarity precluded a response based on the other. We chose the higher-order relation of *monotonic change in size across position.* This relation has the advantage that it can be understood on the basis of perceptual information available to the child (in contrast to some causal or social higher-order relations that may require abstract knowledge). This cross-mapping task is used in Experiment 1, whose results of serve as a baseline level of performance. In Experiment 2, we gave children the relational labels "Daddy/Mommy/Baby" and found a predicted gain in performance. In Experiment 3 we tested other sets of relational labels, and further, tested for long-term effects of labeling.

EXPERIMENT 1

Participants

The participants were 24 3-year-olds, 24 4-year-olds , and 16 5-year-olds.

Procedure

Children were asked to map monotonic change in size between a triad of objects belonging to the experimenter and a triad of objects belonging to the child. A cross-mapping was created by staggering the sizes of the objects, as illustrated in the following diagram in which the objects, represented by numbers, form monotonic change in size from left to right.

E 3 2 1
C 4 3 2

The experimenter and the child sat across from each other with the stimulus sets in two arrays separated by about 6 inches, forming an arc in front of the child. The child closed his eyes and the experimenter hid a sticker underneath one of his toys, as she explained, "I'm going to hide my sticker underneath one of my toys while you watch me. If you watch me carefully, and think about where I hid my sticker, you'll be able to find your sticker underneath one of your toys." She then placed her sticker under a toy in her set and said "If I put my sticker under this toy, where do you think yours is?". The child was then allowed to guess, but kept the sticker only if he found it on his first guess.

Using a relative size rule, if the experimenter chose object 2 in her set the correct choice is the child's object 3 (Notice that the child must resist an object match between the experimenter's object 2 and his object 2.). Thus, object similarity was put in conflict with relational similarity (in the form of monotonic increase) forming a task in which a response based on either similarity type is possible, but only a response based on relational similarity is correct. The children performed 14 cross-mapped trials.

Materials

We designed rich stimulus sets that contained interesting, rich objects that varied along all dimensions, including size, within the two sets (e.g., a red flower in a pot, a wooden house, a green mug, and a race car) and sparse stimulus sets that contained very simple, sparse objects that were identical in all respects but size within the two sets (e.g., clay flower pots). (See Figure 1.).

Based on structure-mapping theory, we predicted that the rich object matches would compete strongly with the relational mapping rule. In contrast, the sparse object matches would be relatively easy to overcome—children would be able to perform the relational mapping despite a common object identity choice. A related prediction was that the children would make significantly more object-identity responses when object richness was high than when object richness was low.

Results and Discussion

The children's correct relational responses revealed both the predicted effect of object richness and the relational shift in analogical performance. The richness effect led the children to produce significantly more relational responses with the sparse stimulus objects (54% for the 3-year-olds, 62% for the 4-year-olds, and 95% for the 5-year-olds) than with the rich stimulus objects (32% for the 3-year-olds, 38% for the 4-year-olds, and 68% for the 5-year-olds), suggesting that the presence of rich, distinctive object matches created a salient alternative to the relational response (at least for young children). In contrast, when sparse objects were used, the object similarity matches were less compelling and therefore less likely to act as a competitive alternative to the relational response. As further evidence for the effect of object richness, the number of object identity errors significantly increased with the use of the rich stimuli (33% for the rich versus 17% for the sparse, collapsed across all three age groups). The relational shift was found in the children's overall performance, with the 5-year-olds performing significantly better than 3- and

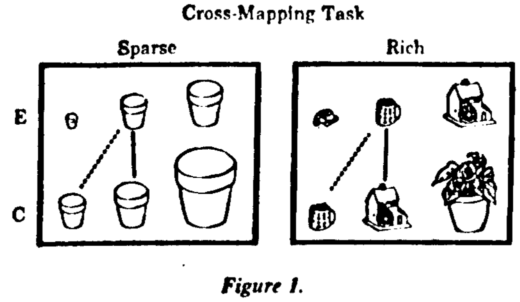


Figure 1.

4-year olds, who achieved above chance performance only with the sparse stimuli.

In Experiment 2 we tested whether a set of relational labels that provide children with an explicit relational structure could help them carry out a relational comparison and mapping. We introduced a group of 3-year-olds to the use of the labels "Daddy/Mommy/Baby" to describe the relationship of monotonic change, and then presented them with the cross-mapping task of Experiment 1. We hypothesized that these labels would provide the children with an explicit framework for the relational system of *monotonic change in size*. If so, then the children's ability to perform a relational mapping with both the rich and the sparse stimuli should improve with the use of these labels. To obtain the maximal effect of labeling, we went to great lengths to ensure that the children were familiar with the relational use of the family labels, training participants with the "Daddy/Mommy/Baby" labels prior to presenting them with the cross-mapping task. We also reminded the children of these labels on each trial during the course of the experiment.

EXPERIMENT 2

Participants

The participants were 24 3-year-olds.

Procedure

Label-training. The label training stimuli were a set of toy penguins and a set of teddy bears, each of four different sizes and with very different markings. Training consisted of two sets of four trials in which the cross-mapping between objects and relations did not hold (bears were mapped to penguins) and two sets of four cross-mapped trials (penguins were mapped to penguins). The following protocol was used for the first eight trials:

"These bears and these penguins are each a family. In your bear family, this is the Daddy (pointing to the larger bear) and this is the Mommy (pointing to the smaller bear). In my penguin family this is the Daddy and this is the Mommy (again pointing appropriately)." When the child successfully labeled the animals in

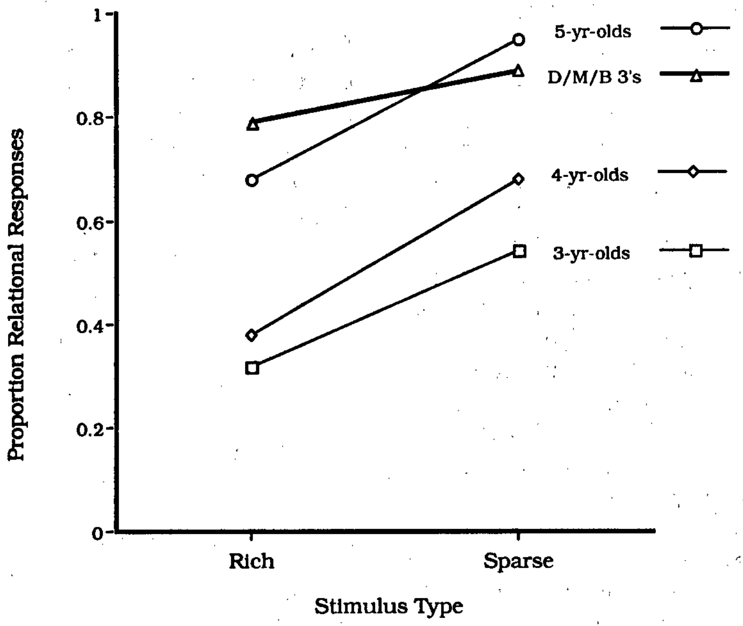


Figure 2.

both sets, the experimenter said "If I put my sticker under my Daddy penguin, your sticker is under your Daddy bear. Look, my sticker is under my Daddy. Where do you think your sticker is?" and the child was allowed to search for the sticker, again only keeping it if he found it on his first guess. After four trials a third, smaller, stuffed animal was added to each set and the labels "Daddy/Mommy/Baby" were applied in the manner described above. The same protocol was adapted for use in the cross-mapped penguin/penguin trials.

Cross-mapping trials. The cross-mapping task from Experiment 1 was used. The children were first asked to label both sets of objects using the family labels (this was repeated every second trial), and then the full-labeling procedure (e.g., "If I put my sticker under my daddy toy, your sticker is under your daddy toy. Look, my sticker is under my daddy, where do you think your sticker is?") was used. The participants each performed 14 sparse trials and 14 rich trials, counterbalanced, although only their performance from the first stimuli type presented was analyzed.

Results and Discussion

The use of the "Daddy/Mommy/Baby" labels did improve young children's ability to make relational comparisons, even in the face of a tempting object choice. When trained to use these labels, 3-year-old children's ability to map relational similarity increased dramatically with both rich and sparse stimulus sets. As can be seen in Figure 2,when "Daddy/Mommy/Baby" was applied to the relation of monotonic increase, the number of relational responses produced by 3-year-olds increased from 54% with the rich stimuli and 32% with the sparse in Experiment 1 to 89% and 79%, respectively, bringing the performance of these participants to the level of performance found in the 5-year-olds. Note, however, that even when the relational labels were used, the effect of object richness was replicated; children produced significantly more relational mappings with sparse objects than with rich objects. Along with the increase in relational mappings, there was a concomitant decrease in the number of object identity errors between Experiments 1 and 2 (from 23% to 8% with sparse and from 43% to 19% with rich).

We propose that "Daddy/Mommy/Baby" helped the young children notice the presence of a familiar higher-order relationship, namely monotonic change, that they may have already represented. Alternatively, the use of the relational labels may have led children to align the two relational systems (the E set and the C set) and derive the common monotonicity structure.

In Experiment 3, we address three further issues. First, we asked whether relational adjectives such as "big/little/tiny" would also enhance children's' ability to perform a relational mapping. Second, we tested for long-term representational change brought about by our use of labels by retesting a sample of 3-year-olds 1-4 months after initial testing. And third, we addressed the possibility that our use of the "Daddy/Mommy/Baby" labels on every trial in Experiment 2 led the children to use the labels as an external crutch—perhaps following the rule "look under the object to which the same label has been applied" without grasping the relationship of monotonic increase in size. To be able to dismiss this possibility we presented children with a small number of full-label trials after which they were given new stimulus sets and asked to perform the cross-mapping without labels being overtly applied.

EXPERIMENT 3

Participants

The participants were 51 3-year-olds, 28 who returned to the laboratory for session 2. The time period between Session 1 and Session 2 varied from 1 month to 4 months.

Materials and Procedure

Session 1. The children were randomly assigned to a labeling condition: no-labels, "Daddy/Mommy/Baby," or "big/little/tiny." Children were given the label-training task used in

Experiment 2, and then eight trials using either the rich or the sparse stimulus sets and the full-label procedure of the previous experiment. After completing the labeled trials the children were shown a new set of stimuli of the same richness type and were given eight trials without labels.

Session 2. To ensure that the testing situation in this later session was as different as possible from the initial session several changes were made; (1) the children were tested using the opposite type of stimuli (i.e., rich or sparse) than was used in their initial testing session; (2) the children were tested in a different testing room, and; (3) a different experimenter performed the experiment. The instructions given to the children were minimal; they were reminded that they had played this game before; "Remember, you played a Daddy/Mommy/Baby game last time. Lets see if you can still play the game." The children were given four practice trials, without labels, using the stuffed penguins and bears. Each child was then presented with eight unlabeled cross-mapping trials, followed by four "reminder" trials in which the full-label procedure was used, and then finally with eight more unlabeled trials.

Results and Discussion

Session 1. Children trained with the relational labels ("Daddy/Mommy/Baby" and "big/little/tiny") produced significantly more relational responses than children in the no-label condition (58% for relational labels and 41% for no-label, collapsed across stimulus complexity and trial type). The effect of richness was replicated; children produced significantly more relational mappings with the sparse stimuli than with the rich stimuli (67% versus 39%, collapsed across label types and trial type). And as in the previous experiments the children produced significantly more object identity errors with the rich than the sparse (37% versus 20%).

Session 2. We first examined children's ability to map monotonic change in the first eight, non-labeled, cross-mapping trials. This data reflects children's ability to apply previously leaned relational structures with minimal prior reminding. The "Daddy/Mommy/Baby" and "big/little/tiny" labels led to more relational responses on these trials than did no labels (62% with the family labels and 54% with the relational adjectives versus 28% with no-labels), suggesting that the children's previous exposure to relational labels had indeed changed their representation of monotonic change. The second aspect of children's relational performance is their performance on the second set of non-labeled trials, after the four trial "reminder" of the relational labels. Overall, children's relational responding increased after being reminded of the relational labels (67% correct with relational labels, versus 45% correct with no labels).

We did not find a significant effect of object richness in this data due to the fact these children were exposed to both types of stimuli across the two experimental sessions. It seems likely that this experience diluted the effect of object richness found in the previous experiments.

GENERAL DISCUSSION

A robust finding in the study of children's analogical abilities is the relational shift (Gentner, 1988; Gentner & Rattermann, 1991; Gentner and Toupin, 1986; Halford, 1993). In Experiment 1 we explicitly tested for the relational shift and found that the presence of a salient object similarity choice disrupted relational mapping in 3- and 4-year-old children, but that 5-year-olds could map relationally despite this conflict, supporting the hypothesized shift from objects to relations in children's analogical reasoning.

In addition to testing for the relational shift, we also made predictions specific to the structure-mapping view of analogy. The predicted effect of object richness, one of the most robust findings in this series of experiments, derives directly from this view. We propose that when performing an analogical mapping, children (and adults) will begin by aligning objects based on common features, and further, that the more salient and numerous the features, the more likely that object matches will win out over relational similarity in the final interpretations (Markman

& Gentner, 1993). In each of our experiments, the presence of a rich object conflict was more detrimental to the ability to perform a relational mapping than the presence of a sparse object conflict. It is worth noting that a similar effect has been found in the performance of adults presented with a cross-mapping task. Markman and Gentner (1993) found that adults will also respond based on object similarity when the number of matching object attributes of the cross-mapped objects is increased.

In the present work young children's susceptibility to rich object matches was due to their incomplete knowledge of monotonic change. We propose that simply using the labels "Daddy/Mommy/Baby," invited children to represent the higher-order relation of monotonic change. We further claim that the addition of this relational knowledge led to a striking improvement—equivalent to that of a 2-year-age gain—in the children's ability to perform relational mappings.

Finally, these experiments show quite forcefully that language, and in particular relational language, can facilitate relational representation. We found that both "Daddy/Mommy/Baby" and "big/little/tiny" led to increased relational responding in our three-year-olds, and that this ability remained several weeks after the initial exposure to these relational labels. The role of language, we suggest, is to provide an invitation to form comparisons and further, to provide an index for stable memory encoding of the newly represented relational structure.

IMPLICATIONS FOR THEORIES OF ANALOGY

The research of Halford and his colleagues (Halford, 1993; Halford, Smith, Dickson, Maybery, Kelly, Bain, & Stewart, 1995) has also found the shift from objects to relations. They propose that an important driver of this shift is changes in cognitive capacity. That is, children show a developmental increase in cognitive capacity that allows them to represent and map increasingly more complex matches. Thus, for example, not until three years should children be able to carry out complex system matches. In contrast, in our account it is domain knowledge that leads to increases in children's analogical abilities.

Neither view of analogy is meant to be exclusive; we acknowledge the role of maturational change in children's cognitive abilities and Halford has consistently noted the role of knowledge. However, our results demonstrate that striking changes in ability can occur over the course of one experimental session, and further that these gains persist after the experimental session is over. It appears that the limits on performance are not in children's capacity to represent and use complex relations, but rather in whether they have as yet represented a given complex relation. These results underscore the point that an increase in relational responding is not evidence, in itself, of maturational gain.

Another prominent theory of analogical reasoning is Goswami's (1993) relational primacy view. Goswami proposes that very young children (3-years-old) can perform an analogy when they have represented the requisite relational structure. While we agree with Goswami that domain knowledge plays a crucial role, we differ in the hypothesized role of object similarity. Goswami has stated that "As long as the relations that the child must map can be represented.... then performing the mapping should present little difficulty, and this should hold true whether the objects to be mapped are similar or different in appearance." (Goswami, 1995, p. 891). However, in our studies there is still a robust effect of object similarity, even when labels have been applied and the children's relational performance is overall very good.

Language and Relations

We have presented the view that labels, and in particular relational labels, invite children to notice and retain patterns of elements; language encourage them to *modify* thought. When applied across a set of cases (or a pair of cases, as here) labels provide children with an invitation to make comparisons, and then provide a system of meanings upon which to base these comparisons (Gentner & Medina, 1997).The results of these labeling studies support this view; 3-year-olds trained with the labels "Daddy/Mommy/Baby" and "big/little/tiny" showed a significant increase in relational responding with a relatively simple linguistic intervention.

The results of these experiments suggest that young children can perform a relational mapping, even in the presence of conflicting object similarity, when familiar labels are used to highlight the appropriate relational structure. The impressive gains in ability after the use of relational labels supports our claim that language provides an invitation for children to modify their thought. Language is not, however, the only path to relational competence. Other manipulations, such as *progressive alignment* in which children are presented with easy *literal similarity* matches prior to difficult analogical matches will also lead to improvement (Kotovsky & Gentner, 1997). Work with primates has also shown that relational labels need not be embedded in a full linguistic system to improve relational responding (Thompson, Oden & Boysen, 1997).

Thus we conclude that one factor in the development of the ability to use relational similarity is the acquisition and use of relational language. Relational language can serve as a catalyst for comparison and alignment of objects and relations, which can, in turn, provide a mechanism for the progression from children's naive thought to the sophisticated, abstract thought of adults.

REFERENCES

Brown, A. L. (1989). Analogical learning and transfer: What develops? In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 369-412). New York: Cambridge University Press.

Crisafi, M. A., & Brown, A. L. (1986). Analogical transfer in very young children: Combining two separately learned solutions to reach a goal. *Child Development, 57*, 953-968.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1990). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 41*, 1-63.

Gentner, D. (1977). Children's performance on a spatial analogies task. *Child Development, 48,* 1034-1039.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155-170.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59,* 47-59.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199-241). London: Cambridge University Press.

Gentner, D. & Markman, A.G. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45-56

Gentner, D. & Medina, J. (1997). Comparison and the development of cognition, Cognition.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman &. J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development.* (pp. 225-277). London: Cambridge University Press.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science, 10*, 277-300

Goswami, U. (1993). *Analogical reasoning in children.* Hillside, NJ: Lawrence Erlbaum.

Goswami, U. (1995). Transitive relational mappings in three- and four-year-olds: The analogy of Goldilocks and the Three Bears *Child Development, 66*, 877-892.

Halford, G. S. (1993). *Children's understanding: The development of mental models.* Hillsdale, NJ: Erlbaum.

Halford, G. S., Wilson, W. H., Guo, J., Gayler, R. W., Wiles, J., & Stewart, J. E. M. (1995). Connectionist implications for processing capacity limitations in analogies. In K. J. Holyoak & J. Barnden

(Eds.), *Advances in connectionist and neural computation theory: Vol. 2. Analogical connections* (pp. 363-415). Norwood, NJ: Ablex.

Holyoak, K. J., & Thagard, P. R. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104,* 427-466.

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9,* 45-75.

Keane, M. T., & Brayshaw, M. (1988). The incremental analogical machine: A computational model of analogy. In D. Sleeman (Ed.), *Third European Working Session on Machine Learning* (pp. 53-62). San Mateo, CA: Kaufmann.

Kotovsky, L, & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67,* 2797-2822.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431-467.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 510-520

Rattermann, M. J., & Gentner, D. (in press). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development.*

Ross, B. H. (1989). Remindings in learning and instruction. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 438-469). New York: Cambridge University Press.

Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (Pan troglodytes) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes, 23,* 31-43.

Waxman, S. R., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development, 1,* 139-156

ANALOGICAL PROBLEM-SOLVING IN PRESCHOOL CHILDREN

M. Bastien-Toniazzo[1], A. Blaye[2], D. Cayol[1]

1 CREPCO U.R.A. 182 du C.N.R.S. (Center for Research in Cognitive Psychology)
2 Laboratoire "Psychologie du Développement" (EA DRED 850) Université de Provence
U.F.R. de Psychologie et Sciences de l'Education 29 av. R. Schuman
13621 AIX-EN-PROVENCE cedex 1
bastien@aixup.univ-aix.fr

INTRODUCTION:

Analogical reasoning is one of the main theme of the psychology of cognitive development. It's probably because it's a constutive developmemental mechanism in that it allows the subject to construct and modify his knowledge in a flexible and adaptive way (Halford, 1993). Our goal is to analyse the conditions favourind analogical problem-solving in 5- to 6 years old children.

Historically, studies on analogical reasoning were first devoted to proportional analogies (a:b::c:d) then to solving new problems (target problems) which refer to known problems (source problems). In the Piagetian point of view, the ability to reason by analogy emerges when the child reaches the formal stage. More recently, researchers showed that younger children did not fail because they are inable to reason by analogy but because their lack of knowledge about objects or causal relations (Goswami & Brown, 1989 ; 1990 and Gentner & Ratterman , 1991). However most of the situations proposed to the children are four terms analogies which only allow to highlight the solution phase process. Other situations as problem analogies require to retrieve the source before to solve the target problem.

In problem solving situations different factors can contribute to improve the transfer of the solution. Brown, Kane & Long (1989) incite their subjects to extract the relational structure common to the two situations. Brown, Kane & Echols (1986) obtain even more efficiency in helping children to bring out the goal structure of the source story. Holyoak, Junn & Billman (1984) show, in their experiment, that when5- to 6- years old children are prompted to map the target on the goal structure they find analogical solution.

On the whole of these researches, the demonstration of analogical reasoning by young children has been obtained in very compeling conditions. In particular, the target problem follows immediatly the source problem and requires an explicit intervention of the adult.

The present research contends that the classical design used in the investigation of analogical problem-solving undermines young children's abilities. The base analog is usually introduced as a story and the probability of extraction of the relational structure is then very low.

In our two experiments, the goal is to show that, when children are able to reach an optimal level of representation of the source without adult explicitation even after a very long delay (one week). The situations proposed to the children are inspire from Holyoak & al (ib.). However, some characteristics have been modified (see later).

EXPERIMENT 1

56 children of kindergarden from a preschool in an underprivileged neighbourhood performed the source problem but one of them being absent for the target problem, the number is reduced to 55 subjects, from 5; 1 years to 5; 10 years (mean age 5; 5 years).

TASK AND MATERIAL

The task consists in deposit in a container placed inside a bigger container pierced of an orifice, objects too large for pass by the orifice which is situated to the diagonal of the container. It is necessary to addition avoid an other container placed to the plumb of the orifice. This task is presented in two analogous problems ("mouse problem", "boys problem") that differentiate by their indices of surface but that present the same diagram of resolution. The diagram of resolution consists therefore in coordinate four schemes: to crumble (E), to roll (R), to join (J) and to make pass (P). Each of these schemes is known children of this age. Note that the plan of solution is here more complex than at Holyoak and al. 1984.) since it requires one more move (to crumble).

"MOUSE PROBLEM"

On a table is posed a canister semi-spherical in plastic thick and transparent. The high of the canister is pierced of a narrow orifice. To the interior and under the orifice is placed a small pot of plastic containing water. A small plate is diametrically opposite the pot (to see an illustration of the equipment, annex 2). The house of the mouse is placed in the face of the child of such manner that the hole is found on the left side. On the table are scattered: a slice of crumb bread and various objects (leaf of supple transparent plastic, cube from woods from 5 cm of side, pencil, ruler of wood of 10 cm).

One tells to the child that it concerns the house of a small smile and that during its absence one wants to put it the bread in its plate; one must pay attention that the bread does not go in the glass of water.

" BOYS PROBLEM "

On a low table is posed a parallelepipedic canister in black cardboard whose anterior face is replaced by a face of transparent plastic. In the face superior of the canister ther is a small orifice. To the interior and under the orifice there is a small red cardboard canister while of the other side is placed a small white cardboard canister (see annex 2). The canister is placed in the face of the child of such manner that the hole is found on the right side. On the table are scattered: a bloc of small cube (Lego 0,5 cm.) encased and various objects (leaf of transparent supple plastic, cube of woods of 5 cm. of side, pencil, ruler of wood of 10 cm).

One tells to the child that it concerns a bedroom where are found the chest to toys of a wicked boy (red canister) and that a nice boy (white canister). One wants to put Lego in the canister of the nice boy and not in that of the wicked boy.

Procedure.

In the source situation, children are confronted with the resolution of a complex problem. Complex means here that the diagram of solution is not known spontaneously by children. However each step of the of solution is familiar.

So as to neutralize possible effects linked to the content of problems, we have alternated the status (source/target) of the two problems: for half of subjects the source problem is the mouse problem and the target problem is the boy problem while for other half the source problem is the boy problem and the target problem, the mouse problem.

In a first time, children, distribute in two groups, are seen individually in an isolated room and are invited to solve one problem. Children of the control group (n= 27) try to solve the task and no assistance is provided them. The task is considered as ended when the child signals it. To children of the experimental group (n= 28), a guidance is brought to each of the of failed step solution (cf. annex 3). A week later, one proposes to all subjects the second problem. To the moment of the target problem, a leaf of alumin-

Experimental Group (n = 28)			Control Group (n = 27)		
success without hint	success after hint	failure	success without hint	success after hint	failure
10 36%	11 39%	7 25%	0	0	27 100%

Table 1. Distribution of performances on the target problem as a function of experimental conditions.

ium replaces the transparent plastic leaf. In case of impass, only one hint is proposed: "have you already told some thing a little equal that could have help you?". The hint concerns therefore only an assistance for the evocation and in no case of stressing the extraction of the common structure to the two problems.

All actions and verbalizations of the subjects are noted by the experimentator.

Results

The order of presentation of the two problems (mouse/ boys and boys/ mouse) having no significant effect, we will not distinguish therefore data of these two modes.

Can-one to speak analogical transfer between source and target?

A first interesting result concerns the total absence of subjects capable to produce, without assistance, the waited solution during the problem source. Thus, alone subjects of the experimental group have been confronted, with the guidance, to the progress of necessary actions for the resolution of the task.

We are going to consider now the distribution of performances during of the target problem (cf. table 1).

The absence of success in the control group confronted with 75% of success in the experimental group suggests that the former are made a real analogy with the solution of the problem source. 10 among 21 subjects (is 48%) having produced the waited solution have succeeded without hint. If, as suggest these data, the success to the target problem necessitates the evocation of the problem source, one can be interested in verbal expressions of this evocation.

Verbal manifestations of the evocation of the source.

The observation of the gap existing between what subjects are capable to complete and the reality of their mental functioning is today largely admitted. If one considers the spontaneous verbal evocation, we note that subjects of the experimental group are more numerous to make this evocation (61% against 41%). The difference is not however significant. Table 3 shows obviousnelys the interpretive problem that puts the verbalization of the evocation: children can evoke verbally without succeeding (cf. the control group) or to succeed without evoking verbally (cf. the experimental group).

Discussion

Performances of the control group witness the fact that the resolution of the type of proposed problem can not be made by the recovery of a strategy of resolution in memory. We have therefore well there a situation demanding an analogical reasoning.

The resolution of the target problem by the majority of subjects of the experimental group puts thus clearly in obviousness the capacity of children of 5-6 years to solve a problem by analogy with a source situation, when the former is itself presented in the form a resolution of problem. Such a result suggests that we have thus created, without directive induction on the part of the adult, conditions allowing subjects to represent the relational structureof the source problem. Note here favorable effects to the transfer in spite the long period, a week, between the resolution of the source and the target.

If one nears these results of these of Holyoak& al (1984) in the condition magical rug, one notes an appreciably equivalent proportion of subjects that find spontaneously the solution (36% here and 30% at Holyoak& al, ib.). The greatest efficiency of the proposed situation here demonstrates in the success after hinting. At Holyoak& al, no child solve the problem in spite of the hint which centers him explicitly on the structure and while the target problem follows immediately the source. In our case, the hinting, simple incentive to the evocation, allows 39% supplementary of subjects to transfer the solution although a week separates situation source and target.

Nevertheless, although that the two situation -source and target- here differ both from the point of view of the presentation of the problem and available resources to solve them (the type of sheet to roll, criticical point for the resolution, differs from a problem to the other), it remains that functional and perceptive similarity between the two problems appear more important than in the condition "magical carpet/ sheet" of Holyoak and al. (ib.).

One could have be tempted to assimilate our situations to those use by Holyoak and al. in condition "magical stick/ stick" condition in which hint seems very efficient since they drove to the success of 100% of subjects. It remains that in this condition, perceptive characteristics of the stick (close to those of the magical stick) are precisely those that suggest its possible function to know, near a too distant container. In our study on the other hand, the undeniable perceptive similarity between a sheet of aluminum and a sheet of plastic does not return to the function of tube. Now, as have underlined it Holyoak and Thagard (1995), the perceptive indication plug is guided by the function.

One of the objectives of the experience 2 is precisely to judge the weight of the perceptive similarity in the strong rate of transfer obtained.

EXPERIMENT 2

Besides Holyoak and al. (1984), many authors (Gentner& Toupin, 1986; Goswami, 1992, for a review) have established that an increase of the perceptive similarity between source and target favors the analogical transfer at children -one observes it also at adults.

One can therefore offer a second interpretive hypothesis to the strong rate of transfer obtained in the experience 1. It would not be for the essential the result of possible productive conditions the representation of the relational structure of the source but well rather than of a similarity such, that it would render very probable the evocation of the source then the mapping between the two situations.

Results of Brown and al.(1986) on analogical situations in which available resources between source and target were strictly identical, suggest that the similarity is not a sufficient condition to the transfer. It appears nevertheless important to test experimentally a such alternative hypothesis to that that we favor.

We have thus confronted the condition studied in the experience 1, "source problem/ target problem" (P-P Condition) on the one hand, to the mode of "classic" presentation (Holyoak and al., ib.), "history source/ target problem" (H-P condition) and, on the other hand, to a condition "history source mimiced by the experimentator/target problem" (condition HM-P). The confrontation of the condition P-P to the alone "classic" H-P condition is still ambiguous on the plan of interpretations. Indeed, besides the fact that in the first, subjects have more the possibility to assimilate the structure, the totality of objects that they have to their disposition to solve the problem source is perceptually identical (except for the sheet to roll) to the available objects during the resolution of the target problem. This is not the case in the H-P condition in which only an illustrated book is read to children. One has therefore there possibly a strictly perceptive indication being able to favor the analogy. The HM-P condition would have to allow to slice between these two hypotheses. It is only in the case where one would obtain a significant superiority of the P-P condition both on H-P and HM-P conditions that we could have reject an interpretation resting exclusively on a perceptive facilitation linked to the objects.

An other manner to push more before the study of the role of the perceptive proximity degree between source and target in the production of a transfer, consists in make it vary experimentally, non from the point of view of available resources for the resolution, but from the point of view of the "environment" (decoration) in which the problem targets is posed.

Finally, although we have noted, during the first experience, a connection between verbal production (in the occurrence, evocation of the source) and analogical transfer, we wished to explore the role of a systematic verbal production asked to subjects to the exit of the source situation. It will concern to ask them to" repeat "the history that one comes to tell them (conditions H-P and ++HM-P) or what one comes to make (condition P-P). Such a verbal restitution task has been employed by Brown and al. (1986). These authors have observed that the restitution in it even had no effect on the then even transfer that it took place immediately before the resolution of the problem targets (since source and target followed immediately). We will tempt an analysis in this senses in observing nevertheless that the importance of the period (a week) that we impose on children would have tender to decrease again the effect of the restitution.

Thus three experimental factors are manipulated in the experience 2 driving to 3* 2* 2= 12 experimental groups: the mode of presentation of the source: P-P/ H-P/ HM-P; the perceptive similarity degree between source and target: close/ far; demand or non of a restitution of the phase source.

Subjects

183 children (mean age: 5; 6 years) have participated in this second experiment. They come from 10 different schools inserted in an standard socio-economic environment. Children of a same classroom are distributed in equivalent manner in each of 12 experimental groups. But as the absence of some of our subjects to the moment of the problem targets, the number of subjects is not strictly identical in each of groups.

Material

The material of the experience 1 has been completed by the material serving to the mode "far" of the factor degree of similarity source/ target (cf. Table 1). An elephant wiht on its back a basket pierced of a small hole, is placed in the low of a mountain, on the external bank of a river materialized between the low of the mountain and the elephant. On the mountain holds a small doll holding a slice of crumb bread. The problem consists in help the doll to put the bread in the basket of the elephant, without that the bread falls to earth or in the river . In a perceptive point of view, the "elephant " problem is different the "boys" problem. On the one hand objects of the situation are not in a closed environment, on the other hand, the size of each object is appreciably greater.

Procedure

The absence of effect, in the experience 1, in the order of presentation of problems, has behaved us to use the same "boys" problem as source for all subjects. The guidance to the solution, in the problem source, is identical to that the experience 1. In condition problem and history resolution mimiced by the experimentatot, various objects are had on the table (pot of yoghourt, trombone, pencil, string of 10 cm., gum, transparent plastic sheet). The solution is shown in moving the sheet of plastic. In condition read history, an illustrated handbook serves as support to the narration (to see annex 5 the text of the history).

For the target problem, one replaces the transparent plastic sheet of the source by a sheet of aluminium. Two degrees of hint are planned if the child does not find spontaneously the solution.Hint 1 suggests the necessity of the evocation:" One you has already told a history that could have help you? ". Hint 2 incites to evoke the history source:" The last week one has told you the history of a nice boy and a wicked boy ". Note that these two relaunchings remain less directives than those Holyoak & al (ib.) where authors evoked first the history that came to be told then cneter the attention on what had made the genious.

Results

Results of experiment 2 are presented in the table 2. Only subjects having strictly succeeded to solve the problem targets in chaining 4 schemes Crumble, Roll, Join and Pass are accounted (some children have rolled the sheet of aluminium then have failed).

In lines, the conditions of presentation of the source problem ; in columns, the performances on the target problem.

The examination of the number of success to the target problem in the three conditions reveals, in accordance with the hypothesis that we privilege, a significant effect of the factor condition of presentation of the source (Chi squared= 29,27; p<.0001). Especially, the condition P-P establishes more efficient than each the two other conditions (P-P vs HM-P: Chi squared= 4,661; p=.0309; P-P vs H-P: Chi squared= 10,561; p=.0012). One observes more, a greatest number of successes in the condition where the history source is mimiced that in that where it is read (HM-P vs H: Chi squared= 9,735; p=. 0018). The hierarchy between the three mode presentation of the source is found unchanged when one considers proportions of success without hint.

Globally, the proportion of success to the target problem does not reveal difference between the two perceptive proximity degrees (mouse vs elephant) between source and target: this proportion is 59% in close proximity condition and 53% in condition of weaker proximity. Taking in counts the moment of the success, one notes nevertheless a number of success, without hint, significativly more high when the proximity between source and target is the close (Chi squared= 5,118; p=.02). This superiority results in fact from the alone condition P-P.

The restitution asked at the end of the phase source produces no global significant effect on performances. One observes however in the condition H-P a tendency to what subjects having had to restitute the history are more numerous to solve the problem targets (corrected squared Chi= 3,254; p=.0712).This result could have possibly be found reservations according to the quality of restitutions. We have distinguished three types of restitution : these that clarify clearly the totality of the solution; these that mention only some schemes but the critic one "to roll" and finally these that mention only some elements of material.

These results does not allow to reveal a value forecasts the quality of the restitution. One

		target problem "mouse"				target problem "elephant"			
		success without hint	success after hint	failure	*number of subj.*	success without hint	success after hint	failure	*number of subj.*
P	rest	9 56%	5 31%	2 13%	*16*	5 31%	6 38%	5 31%	*16*
	non-rest	11 85%	0	2 15%	*13*	6 38%	6 38%	4 25%	*16*
HM	rest	6 40%	3 20%	6 40%	*15*	5 31%	6 38%	5 31%	*16*
	non-rest	3 19%	6 38%	7 44%	*16*	4 27%	3 20%	8 53%	*15*
H	rest	3 18%	5 29%	9 53%	*17*	0	5 36%	9 64%	*14*
	non-rest	2 13%	1 7%	12 80%	*15*	0	2 14%	12 86%	*14*

table 2. Distribution of performances on the target problem as a function of the conditions of presentation of the source problem and the degree of perceptual similarity between base and target.

observes only a tendency to what the success without hint is less probable at subjects having restored only elements of material. In the condition H-P where the restitution seemed to favor the performance, one observes that to an alone exception near, all subjects solving the problem targets are these that have verbally extracted the structure of the problem source during the restitution.

Furthermore, the quality of restitutions does not allow to distinguish subjects according to the condition of processing of the source.

DISCUSSION OF EXPERIMENT 2

This second experiment reinforces the hypothesis according to whether it is well the quality of processing that allows the activity of problem resolution in source that would be responsible the good performances observed on the target during the experience 1. In accordance with a classic data of the literature, one observes however a share of facilitation linked to the perceptive proximity degree. More precisely, it appears here that a greatest proximity favors the evocation of the source since subjects are then less numerous to to have need hints inciting them to evoke the source.

The global effect absence of the restitution establishes true to already reminded results of Brown and al. (1986). It is interesting to note that one observes an effect in the condition by less favorable hypothesis to the centration of subjects on the structure of the problem source (condition H-P). It would seem therefore that the task of restitution could contribute to this focalisation subjects while the read history did not the incite there. This interpretation is reinforced by the observation according to whether, the quasi totality of children making analogy in this condition are these that have clarified completely the structure of the problem.

In the two other conditions of presentation of the source, one does not find differences linked to the quality of the restitution. This observation could have partly result from the long period between the moment of the restitution and the resolution of the target (a week). But, in our opinion, more fundamentally, it puts a time news the question of the verbal data reliability as reflection of the processing undertaken by subjects.

GENERAL DISCUSSION

Since the middle of 80s, date of the pilot study of Holyoak and al. (ib.), the vision of psychologists on capacities of kindergarrners to solve problems by analogy establishes appreciably more optimistic (cf Holyoak & Thagard 1995).

The totality of presented data here demonstrates the capacity of children of 5-6 years to solve problems by analogy. This demonstration establishes particularly convincing and this to more of a title.

On the one hand, the strong rate of transfer obtained in the condition problem-problem, it has been in spite an extraordinarily long period between situation source and problem targets. To our knowledge, no study had tempted to impose such a period to children, privileging situations where source and target follow immediately.

On the other hand, as we have reminded it in introduction, factors studied until here for their efficiency in the analogical transfer of the young children imply quasi-systematically a very explicit intervention of the adult driving children to extract the relational structure in the situation source. One will retain that here hints did not go beyond an assistance with the evocation of the source. This difference is obviously fundamental to the extent of, as underline Holyoak & Thagard (1995), the period of 4 to 6 years constitutes precisely a phase from transition in the course of which elaborates the capacity of mapping of systems of relationships and only relationships of first order. In brought studies here, the mapping has never been explicitly suggested. More, the superiority of the processing of the source in resolution of problem as compared to a history mimiced and told source allows to reject the hypothesis of an help with themapping provided by the alone perceptive similarity between resources proposed in source and in target.

Thus, the appropriation of the source in the form of a resolution of problem establishes a sufficient condition to allow an analogical transfer to a majority of children of 5-6 years, even in the situation of lesser perceptive proximity between the two problems. The absence of difference concerning the verbal restitution quality between the different conditions of processing of the source does not allow to exhibit a better explicitation of the structure of purpose of the source at subjects placed in Problem-Problem. This point of view, we have obtained a direct validation of the hypothesis according to whether it is because it allows a focalisation on the relational structure that the resolution of the source is a favorable condition to the transfer. Nevertheless, the verbal restitution activity demands a level of explicitation (in the sense of Karmiloff-Smith, 1992) of the representation that has elaborate the subject that does not seem a necessary condition for the possibility of a mapping with the problem targets.

As has proposed it Karmiloff-Smith, to solve a problem by analogy requires effectively a "representational redescription" of the source so as to to adapt it to the resolution of the target. We know, according to this author, that a preliminary condition to the possibility of redescription is the "comportemental mastery" of the initial situation. Studies presented in this article reinforce this thesis and suggest that one of the best ways to acquire this mastery consists precisely in place subjects in situation of guided problem resolution of the source.

REFERENCES

Brown, A. L., Kane, M. J. & Echols, C. H. (1986) Young children's mental models determine analogical tranfer across problems with a common goal structure, *Cognitive Development*, 1, 103-121.

Brown, A.L., Kane, M. J. & Long, C. (1989) Analogical transfer in young children : analogies as tools for communication and exposition, *Applied Cognitive Psychology*, 3, 275-293.

Brown, A. L. & Smiley, S.S. (1977) Rating the importance of structural units of prose passages : problem of metacognitive developement, *Child Developement*, 48, 1-8.

Gentner, D., & Ratterman, M. J. (1991) Language and the career of similiraty, in S. A. Geldman & J. P. Byrnes (Eds), *Perspectives on language and thought : Interrelations in development*. Cambridge : Cambridge Univesrity Press.

Gentner, D. & Toupin, C. (1986) Systematicity and surface similatity in the development of analogy, *Cognitive Science*, 10, 277-300.

Goswami, U. (1992) *Analogical reasoning in children. Essays in developmental psychology*, Hillsdale, N.J. : Erlbaum.

Goswami, U. & Brown, A. L. (1990) Higher-order structure and relational reasoning : contrasting analogical and thematic relations, *Cognition*, 36, 207-226.

Goswami, U. & Brown, A. L. (1989) Melting chocolate and melting snowmen : analogical reasoning and causal relations, *Cognition*, 35, 69-95.

Halford, G.S. (1993) *Children's understanding : The development of mental models*. Hillsdale, N.J. : Erlbaum.

Holyoak, G. S. & Taghart, P. (1995) *Mental leaps : Analogy in creative thought*. Cambridge, Massachusetts : M.I.T. Press.

Holyoak, K. J. & Hoh, K. (1987) Surface and structural similarity in analogical transfer, *Memory and Cognition*, 15, 332-340.

Holyoak, K. J., Junn, E. N. & Billman, D. O. (1984) Development of analogical problem-solving skill, *Child development*, 55, 2042-2055.

Karmiloff-Smith, A. (1992) *Beyond modularity : a developmental perspective on cognitive science*, Cambridge, Massachussets : M.I.T. Press.

CONCERNING THE ROLE OF ANALOGY IN METAPHOR PROCESSING

John A. Barnden
School of Computer Science, The University of Birmingham
Birmingham, B15 2TT, United Kingdom
J. A. Barnden@cs.bham.ac.uk

ABSTRACT

In understanding a metaphorical utterance, there is the question of how to use the analogical mapping (if any) associated with the metaphor, once this mapping is known. It is usually assumed that one should translate the situation literally depicted by the utterance into terms of the target domain, and that this requires extending the mapping to source items and structure that are not yet mapped by the analogy. However, this paper argues that it is mistake to think that such extension must generally be done. This mistake arises from an unworkable assumption that metaphorical utterances must generally be assigned meanings other than their literal ones. Instead, the paper advocates an approach that treats a literal meaning as a basis for an indefinite amount of within-source inference connecting with the existing analogical mapping, but does not seek to extend it. This approach has been implemented, in a reasoning system called ATT-Meta.

INTRODUCTION AND APPROACH

Consider an utterance of the sentence "Those two ideas were in different corners of John's mind." This rests upon the metaphors of MIND AS PHYSICAL SPACE and IDEAS AS PHYSICAL OBJECTS. In this paper, metaphors are conceptual views, not utterances or parts of utterances (cf. Lakoff 1993). Rather, the utterance is merely one possible linguistic "manifestation" of the metaphor(s). I assume that "corners" has as one of its primary literal meanings the corners of a room, and that therefore John's mind is being metaphorically viewed as being or containing a room. Let's suppose that the addressee (hearer, reader) of the utterance is familiar with the two metaphors mentioned, but has not before encountered the idea of something being in a "corner" of someone's mind. How is the addressee to make sense of the sentence?

Let's assume that the addressee takes the analogy underlying MIND AS PHYSICAL SPACE to include the following: a mind corresponds to a bounded or unbounded physical region; and mental entities or events correspond to entities or events located in that region. I will assume that for the addressee the analogy underlying IDEAS AS PHYSICAL OBJECTS the includes the following: someone's ideas (special case of mental objects) correspond to some physical objects; and *inferential interaction between mental objects or events corresponds to physical interaction* of the corresponding physical entities. But let's suppose that the analogies say nothing specifically about what it is about the mind that "corners" correspond to, or how being in "corners" could possibly be significant for mental entities.

Consider an utterance that manifests some metaphors. Let M be a subset of these metaphors. A "source-based" meaning of the utterance, with respect to M, is one that arises from treating the metaphors in M as objectively true views. That is, in our corners example, the source-based meaning with respect to both the MIND AS PHYSICAL SPACE and IDEAS AS PHYSICAL OBJECTS metaphors casts the mentioned ideas as literally being physically situated in physical corners that are literally physically inside the person's mind. By contrast, a "target-based" meaning of the utterance, with respect to M, would cash out the metaphors in M in terms of of their targets. To switch examples, the target-based meaning of "The idea was in John's mind", with respect to the MIND AS PHYSICAL SPACE metaphor, could be that John was considering the idea in some way. In cases where an utterance manifests only one metaphor, source-based and target-based meanings can be called "literal" and "metaphorical" meanings respectively.

Many metaphor theorists appear to assume that (A) the goal of metaphorical-utterance processing is to construct a representation of a target-based meaning. (For instance, much psychological research on metaphor makes reference to the "metaphorical" — i.e., target-based — meaning, tacitly assuming that it exists and needs to be determined.) It then appears obvious that (B) this representation should contain elements that correspond to the major source-domain elements that appear in the utterance. So, in our example, the representation should involve aspects of John's mind that are being viewed as corners, in some suitable extension of the analogy involved in MIND AS PHYSICAL SPACE. Also, the property of something being in a corner would have to map over as well, to some property of ideas. Since the analogies behind the two metaphors manifested by the corners utterance have nothing to say directly about corners, we have a case of unmapped source-domain entities and associated structure being mapped (transferred) to the target domain.

Indeed, analogy processing is classically divided up into retrieval, matching, transfer, evaluation and so forth, with transfer being central when an analogy is used creatively, such as when an utterance is a novel manifestation of a familiar metaphor. The purpose of the transfer phase is typically to transfer (in adapted form) as much unmapped structure as possible from the source to the target. A common variant on this idea is to do transfer in a goal-directed way, where the goals arise from reasoning tasks in the target domain. Nevertheless the idea is still that of mapping unmapped structure.

In Martin's (1990) work, the system attempts to map so-far unmapped source items to the target. The SAPPER system (Veale & Keane 1997) is prolific in its attempts to extend mappings. Grady (1997), in discussing the THEORIES AS BUILDINGS metaphor, points out that things such as French windows or tenants in a building have no natural correspondence with anything in theories, and implies that in order to make sense of sentences mentioning the French windows or tenants of a theory the understander must use additional metaphors to map the French windows or whatever. He therefore seems to subscribe to forms of (A) and (B). On the other hand, as an exception to the present comments, Hobbs (1990) appears not subscribe to (A) and (B) in dealing with familiar metaphor; indeed, the approach in this paper has some strong similarities to his. (Of course, if the task is to deal with entirely novel *metaphors* in an utterance, as opposed to novel manifestations of familiar metaphors, then unmapped structure must perforce be mapped. This paper is not about dealing with novel metaphors.)

Going back to our "corners" example, my claim is that it is extremely hard, if not impossible, to come up with a convincing mapping from physical corners to aspects of minds, even though the utterance is readily understandable. We simply do not know enough, whether commonsensically or scientifically, about how the mind works. (And I hope that it is fairly obvious from what follows that if there isn't anything in John's mind that can confidently be cast as "corners" then there's no point, from the point of view of discourse understanding, *imposing* on John's mind the stipulation that it contain something corresponding to corners.) The claim I am making is just an expression of the notorious unparaphrasability of many metaphorical utterances. However, the notoriety has not gained sufficient respect among researchers who actually devise metaphor-processing schemes.

As a variant on the corners utterance, one could say "Those ideas were in different recesses of John's mind." If one attempts to map corners, one should presumably also attempt to map recesses. The utterances have much the same effect but are subtly different in their connotations. The recesses sentence conveys more strongly that the ideas were "hidden away" or inaccessible. But do we know enough about the mind to say whether corners and recesses themselves should map to subtly different aspects of John's mind, and to say what those aspects are? (Answer: no.) Isn't it rather that this new example and the old one convey that the ideas in question were in somewhat subordinate, hidden or inaccessible positions in John's mind, where moreover those positions were relatively inaccessible or distant from each other, so that physical interaction between the ideas was unlikely? If so, then the mapping mentioned above between physical interaction and inferential interaction comes into play, to generate the connotation that the ideas (probably) did not inferentially interact.

The process of going from the ideas being in corners or recesses to the hypothesis that the ideas do not physically interact is one of *within-source* reasoning (or *within-vehicle* reasoning as I have called it elsewhere) conducted on the basis of the *source-based* meaning of the utterance. The amount of within-source reasoning needed to meet a source element such as physical interaction that is mapped over to the target is not in principle limited.

But in special cases no within-source inference may be needed at all — if the source-based meaning is one that can be directly mapped by the existing analogical mapping. For example, the utterance might be "these ideas were in John's mind," which might be immediately mappable by the analogy to the proposition that John was considering the ideas. This proposition can then be called the target-based meaning. But in general the advocated approach does not lead to anything that can straightforwardly be called the target-based meaning. That is, the approach is *semantically agnostic* with respect to target-based meanings. It is akin to but less extreme than the viewpoint of Davidson (1979). Davidson likewise puts great weight on connotations drawn from the source-based meaning by pragmatic processes. However, unlike the case in Davidson there is no objection in the present paper's approach to someone taking the terminological stance that, say, any set of inferences that happens to be drawn from the source-based meaning of an utterance in context constitutes the target-based meaning (in context) for the utterance.

Fortunately, to accompany the claim that target-based meanings (in a more traditional sense than just imagined) often cannot be found, I claim that it is often or perhaps even typically not necessary to find them in the first place. What is important is for any given metaphorical sentence to contribute information to the overall discourse. Whether it does so through the medium of target-based meanings or in the looser way described above is a secondary issue. Notice that, in practice, a sentence like "Those two ideas were in different corners of John's mind" will be in some context where it *matters* that the ideas in question are in different corners. For instance, the discourse might be of the form "... John didn't see the consequences of [some ideas]. They were in different corners of his mind...." If the understander is predisposed to seek coherence relationships between sentences, and tries the idea that the second is an explanation for the first, then understander will be primed to investigate John's drawing of inferences from the ideas. Thus, the above within-source inferences could be constructed backwards, in a goal directed sense, meeting up eventually with the source-based meaning.

To go back to the discussion of THEORIES AS BUILDINGS, suppose that someone says "Mary overhauled the theory from the plumbing to the chimney-stacks." This surely connotes that Mary very thoroughly overhauled the theory. We can get this connotation without having to worry at all about what features of theories correspond to plumbing and chimney-stacks. It is a plausible conjecture that those items are only mentioned by the speaker to emphasize that the overhaul is thorough. So, it is simply that within-source inferencing is needed to infer that the physical overhaul is thorough. Assuming then that, under the metaphor, physical overhauling maps to large-scale modification of the theory, and that thoroughness of physical overhaul maps to thoroughness of that modification, the connotation mentioned above can be inferred.

The advocated approach refrains from assuming that the point of metaphor is for patterns of inference in the source domain to be mapped to or imposed upon the target domain, as is assumed in much writing on metaphor (incl. Black, 1979; Lakoff & Turner, 1989). This is not to say that such mapping or imposition cannot be done or should not ever be done, but just that for most cases of understanding novel manifestations of metaphors, it is not necessary (and may not be possible) to map inference patterns that are not already mapped by the analogy underlying the metaphor. (Of course, that existing analogy will typically map some inference patterns over to the target.) The claim is that often, or even normally, the products of within-source inference are what's important, not their pattern. But it is certainly possible for novel manifestations of a metaphor to require unmapped source entities and structure to be mapped to the target. For instance, if someone describes something as being te the chimney-stack of the theory" then obviously some target correspondence for chimney-stacks must be found.

The rest of the paper is mainly about how the advocated approach is fleshed out in the ATT-Meta system. The remaining sections of the paper are as follows: a section on the main type of metaphorical utterance considered in the research; a section very briefly sketching ATT-Meta's basic reasoning facilities, irrespective of metaphor; a section describing ATT-Meta's metaphorical reasoning; a section on various types of uncertainty handled in ATT-Meta's metaphorical reasoning, and on an observation that metaphor-based inferences should often override target information; and a section containing final remarks.

METAPHORS HANDLED BY ATT-META

The ATT-Meta system does not currently deal with novel *metaphors* — rather, it has pre-given knowledge of a specific set of metaphors, including MIND AS PHYSICAL SPACE and IDEAS AS PHYSICAL OBJECTS. But it is specifically designed to handle novel *manifestations* of those metaphors. Its knowledge of a metaphor consists mostly of a relatively small set of very general "conversion rules" that map between the source and target domains. These encapsulate what the system knows about the analogy behind the metaphor. The degree of novelty that the system can handle in a manifestation of a metaphor is limited only by the amount of knowledge it has about the source domain and by the generality of the conversion rules.

The ATT-Meta research has concentrated on metaphors for mental states, although the principles and algorithms implemented are not restricted to or specialized for such metaphors. Mundane types of discourse, such as ordinary conversations and newspaper articles, often use metaphor in talking about mental states/processes of agents. Indeed, as with many abstract topics, as soon as anything at all subtle or complex needs to be said, metaphor is practically essential. There are many mental-state metaphors apart from the two mentioned already. Some are as follows: COGNITION AS VISION, as when understanding, realization, knowledge, etc. is cast as vision, as in "His view of the problem was blurred;" IDEAS AS INTERNAL UTTERANCES, which is manifested when a person's thoughts are described as internal speech or writing (internal speech is not *literally* speech), as in "He said to himself that he ought to stay at home and work;" and MIND PARTS AS PERSONS, under which a person's mind is cast as containing several sub-agents with their own thoughts, emotions, etc., as in "Part of him was convinced that he should go to the party." Many real-discourse examples of mental-state metaphor can be found in a databank at **http://www.cs.bham.ac.uk/jab/ATT-Meta/Databank**.

As well as being able to reason metaphorically about agents' beliefs and reasoning, ATT-Meta has general, non-metaphor-related facilities for reasoning about agents' beliefs and reasoning. These facilities are beyond the scope of the present paper, but are described in Barnden (1998) (see also Barnden, to appear, and Barnden, in press; and see Barnden *et al.* 1994 for an early version).

ATT-META'S BASIC REASONING

ATT-Meta is merely a reasoning system, and does not deal with natural language input directly. Rather, a user supplies hand-coded logic formulae that are intended to couch, albeit simplistically, the source-based meaning of small discourse chunks (typically two or three sentences).

ATT-Meta is rule-based, and manipulates hypotheses (facts, conclusions or goals), represented as expressions in a situation-based or episode-based first-order logic somewhat akin to that of Hobbs (1990). At any time, any particular hypothesis H is tagged with a certainty level, one of certain, presumed, suggested, possible or certainly-not. The last just means that the negation of H is certain. Possible just means that the negation of H is not certain but no evidence has yet been found for H itself. Presumed means that H is a *default*: i.e., it is taken as a working assumption, pending further evidence. Suggested means that there is evidence for the hypothesis, but it is not (yet) strong enough to enable H to be a working assumption.

ATT-Meta applies its rules in a backchaining style. It is given a reasoning goal, and uses rules backwards to generate supporting subgoals. When a rule application supports a hypothesis, it supplies a level of certainty to the hypothesis, calculated as the minimum of the rule's own certainty level and the levels picked up from the hypotheses satisfying the rule's condition part. When several rules support a hypothesis, the maximum of their certainty contributions is taken.

When both a hypothesis H and its negation –H are supported to level presumed, *conflict-resolution* takes place. The system attempts to see whether one hypothesis has more specific evidence than the other. If a hypothesis is more specifically supported than its negation, it stays presumed and the negation is downgraded to suggested. If neither hypothesis wins, both are downgraded to suggested. Under certain conditions, one way for a hypothesis to be more specifically supported than its negation is for it to be supported (directly or indirectly) by a proper superset of the facts supporting the negation. Inter-derivability relationships between hypotheses appearing in the support networks are also used in specificity comparison.

This paper will not display ATT-Meta's formal representations and formal rule formats (which are in turn represented as Quintus Prolog expressions), and will use English glosses instead. These glosses may use the past tense to match the tense of English example sentences, but this is just for readability, and ATT-Meta currently has no working treatment of time. Detail on the representational style is in Barnden *et al.* (1994), and considerable detail on ATT-Meta's general reasoning framework (and belief reasoning) can be found in Barnden (1998). As explained in Barnden (1998), ATT-Meta's algorithms for metaphor-based reasoning are almost identical to those for belief reasoning.

ATT-META'S METAPHORICAL REASONING

We will continue to consider the corners sentence, and to assume that it has the following connotation:

Connotation: *The mentioned ideas (as mental entities of John's) did not inferentially interact.*

ATT-Meta's approach to deriving such a connotation involves *source-based pretence* (or *literal pretence* as I have called it elsewhere). A source-meaning representation for the metaphorical input utterance is given to the system, and the system *pretends* that this representation, however ridiculous it is in reality, is true. Within the context of this pretence, the system can do any amount of within-source reasoning (reasoning that arises from its knowledge of the source domains of the metaphors involved) using the source-based meaning. In our example, it can use knowledge about mundane physical objects, rooms and corners. The key point is that this within-pretence reasoning from the source-based meaning of the utterance link up with the analogical mappings involved in the metaphor. In the present case, as explained in the Introduction, the relevant mapping is that from (lack of) physical interaction to (lack of) inferential interaction.

That mapping is itself of a very fundamental, general nature, and does not, for instance, rely on the notion of corners or rooms. *Any* process of within-source inference that linked up with physical interaction could lead to conclusions that ideas did or did not inferentially interact. There is no need at all for ATT-Meta to have any knowledge of how corners or rooms match to aspects of the mind.

ATT-Meta proceeds as follows in dealing with the hand-constructed logical input corresponding to the corners sentence. This input is paraphrased as source-based premises (L1) to (L6) below. The system uses a computational environment called a *metaphorical pretence cocoon* to hold those premises and the within-pretence reasoning. The following shows hypotheses that are placed inside and outside the cocoon:

Inside the Cocoon

((L1)) Idea1 is in corner1.
((L2)) Idea2 is in corner2.
((L3)) Corner1 is a corner of John's mind.
((L4)) Corner2 is a corner of John's mind.
((L5)) Corner1 and Corner2 are distinct.
((L6)) John's mind is a room.

Outside the Cocoon

((SL.i)) I (the system) am pretending that (L.i) holds.
(for i from 1 to 6)

Actually, to include (L6) is an over-simplification, because of course containers other than rooms have corners; also, that John's mind is presumably a container can follow by within-source reasoning from the fact that it has corners, so there is no real need to include something like (L6) from the start.

Given that the system knows that a room is a physical space (or, more precisely, has a physical space as a part), it can infer within the pretence cocoon that John's mind is a physical space. It can also infer that idea1 and idea2 are presumably physical objects, because normally only physical objects are in physical corners. Thus, it is at this point that the system in essence realizes that the utterance manifests the metaphors of MIND AS PHYSICAL SPACE and IDEAS AS PHYSICAL OBJECTS.

As usual, the system is given a reasoning goal, say

((G1)) John believes P.

Suppose that P is some proposition that can be inferred from the ideas mentioned in the utterance (idea1 and idea2). By a process explained in Barnden (1998), (G1) leads to the task of seeing whether

((G2)) idea1 and idea2
inferentially interact.

(cf. the Connotation above). Now, the analogical mapping between physical interaction and inferential interaction appears in the system as a small collection of "conversion" rules, converting between metaphorical and non-metaphorical terms. One such rule can be paraphrased as

Conversion Rule CONV

IF I (the system) am pretending that ideas J and K of agent X are physical objects

AND I am pretending that J and K do not physically interact

THEN **[presumed]** J and K do not inferentially interact.

The presumed is the rule's own certainty qualifier, and serves as an upper bound on the certainty that can be attached to the rule's conclusion by virtue of an application of the rule. In backwards application to the negation of (G2), which is investigated along with (G2) itself, CONV leads to the creation of the subgoal

((G3)) I (the system) am pretending that J and K do not physically interact.

All the goals so far mentioned are *outside* the metaphorical pretence cocoon, but (G3) is automatically accompanied by the subgoal

((G4) J and K do not physically interact

within the cocoon. Now, as part of the system's knowledge about physical objects and space, there is the rule:

IF two physical objects are physically separated

THEN **[presumed]** they do not physically interact.

Within the cocoon the system therefore gets the reasoning subgoal that J and K are physically separated. With the aid of rules about corners, things in corners, and separation, the system can establish this subgoal to certainty level presumed on the basis of the facts (L1) to (L5) in the cocoon. (These facts are certain.) As a result, (G4), (G3), (G2) and (G1) attain level presumed. Notice carefully that the inferencing supporting (G4) is entirely "within-source": it is merely uses commonsense knowledge about mundane physical objects and physical space.

A hypothesis like "I (the system) am pretending that P" is called a *pretence hypothesis*. For each such formula outside the cocoon, P appears inside the cocoon, and conversely. The hypotheses within the cocoon are noted as being within the cocoon by being tagged with the system's name for the cocoon. Such tags are passed around by reasoning rules, so that rule applications on hypotheses within the cocoon lead only to within-cocoon hypotheses. But the tags do not otherwise affect rule application. Thus, application of a rule within a cocoon is virtually identical to application outside the cocoon. And, currently, all rules available for the system's reasoning outside cocoons can also be used within cocoons.

UNCERTAINTY IN METAPHOR

Because reasoning within a cocoon uses the same algorithms as that outside, uncertainty is handled within a pretence cocoon just as it is outside. Partly as a result of this, ATT-Meta includes the following three types of uncertainty handling in its metaphor-based reasoning.

(U1) Given an utterance, it is often not certain what particular metaphors or variants of them are manifested. Correspondingly, ATT-Meta may merely have presumed, for instance, as a tentative level of certainty for pretence premises like the (SL.i) above (even though the L.i themselves are certain). This hypothesis is then potentially subject to defeat.

(U2) Conversion rules like CONV are merely default rules (i.e. their strength is presumed). There can be evidence against the conclusion of the rule. Whether the conclusion survives as a default (presumed) hypothesis depends on the relative specificity of the evidence for and against the conclusion. Thus, whether a piece of metaphorical reasoning overrides or or is overridden by other lines of reasoning about the target is matter of the peculiarities of the case at hand. It is incorrect to think that the target to always have the upper hand, because its own information may itself be uncertain. It must be realized that, just as with non-metaphorical utterances, a metaphorical utterance can express an exception to some situation that would *normally* apply in the target domain. To say "The company nursed its competitor back to health" contradicts default knowledge that companies do not normally help their competitors, and should override that knowledge.

(U3) Knowledge about the source domain of the metaphor is itself generally uncertain. Correspondingly, in ATT-Meta the hypotheses and reasoning within the cocoon are usually uncertain. For instance, it is not certain that physical objects do not interact because they are physically separated, a default we used in the corners example. Thus, conversion rules map within-cocoon information that is usually already uncertain, so for this reason alone their results are generally uncertain.

Because there is uncertain reasoning both within and outside the cocoon, special complications arise for conflict resolution. A particular complication is that the pretence cocoon can contain as a fact any fact sitting outside. This importation of facts is needed because arbitrary information about, say, physical objects may be needed in a pretence cocoon used for a metaphor like IDEAS AS PHYSICAL OBJECTS. Also, non-source-domain rules can be used within the cocoon. But the imported facts and the non-source rules may support something that conflicts with conclusions drawn from the special metaphorical facts inserted into the cocoon at the start (like the L.i facts in the corners example). However, the system adopts the heuristic that metaphorical facts like the L.i supply added specificity. Therefore, ATT-Meta proceeds as follows: within a metaphorical pretence cocoon, specificity-comparison is first attempted in a mode where all reasoning lines partially dependent on imported facts are thrown away. Only if this does not yield a winner are those lines restored, and specificity reassessed. This means that imported facts are downplayed in their effects.

Because of the multiple environments in which reasoning is done (namely: the system's top-level environment; pretence cocoons; similar environments used for simulating agents reasoning) and because pretence cocoons and agent-reasoning simulation environments can be nested within each other to arbitrary depth, conflict-resolution in ATT-Meta is a complex matter that has to proceed down through layers of nesting in an appropriate way. This matter is addressed in Barnden (1998) in the case of reasoning-simulation environments, but the same process also applies when pretence cocoons are thrown in (except for the added specificity provision in the previous paragraph, and a reflection of it into outer layers).

FINAL REMARKS

ATT-Meta's metaphorical pretence processing appears to provide a partial implementation of the "conceptual blending" notion of Turner & Fauconnier (1995). Metaphorical pretence cocoons can contain a mixture of pretence-based and non-pretence-based reasoning, because of the fact importation mentioned in a previous section, and because non-source-domain rules can be used within a cocoon.

Some psychological research suggests the people may not construct source-based meanings for metaphorical utterances, at least if the utterances are in an appropriate context and are of a familiar nature. Although ATT-Meta is not meant to be a psychological model, it is worth noting that its use of source-based meanings does not conflict with the psychological research. This is explained in Barnden (to appear).

A near-future topic for research on ATT-Meta is mixed metaphor. We have seen the mixing of MIND AS PHYSICAL SPACE and IDEAS AS PHYSICAL OBJECTS in this paper, but more conflictful mixing is of interest. Also, this mixing is "parallel" in that a single target is directly illuminated by two metaphors. "Serial" mixing (i.e. chaining), when A is viewed as B and B is viewed as C, can be handled in ATT-Meta by nesting a pretence cocoon for B-as-C inside one for A-as-B. Not much experiment has yet been done on this with ATT-Meta.

ACKNOWLEDGMENT

The research was supported in part by grant number IRI-9101354 from the National Science Foundation (USA).

REFERENCES

Barnden, J. A. (1998). Uncertain reasoning about agents' beliefs and reasoning. Technical Report CSRP-98-11, School of Computer Science, The University of Birmingham, U. K. Invited submission to *Artificial Intelligence and Law*.

Barnden, J. A. (in press). An AI system for metaphorical reasoning about mental states in discourse. In Koenig, J-P. (Ed.), *Discourse and Cognition: Bridging the Gap*. Cambridge University Press.

Barnden, J. A. (to appear). Combining uncertain belief reasoning and uncertain metaphor-based reasoning. To appear in *Procs. Twentieth Annual Meeting of the Cognitive Science Society*, University of Wisconsin-Madison, August 1–4, 1998.

Barnden, J. A., Helmreich, S., Iverson, E. & Stein, G. C. (1994). An integrated implementation of simulative, uncertain and metaphorical reasoning about mental states. In J. Doyle, E. Sandewall & P. Torasso (Eds), *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference*. Morgan Kaufmann.

Black, M. (1979). More about metaphor. In A. Ortony (Ed.), *Metaphor and Thought*, pp.19–43. Cambridge University Press.

Davidson, D. (1979). What metaphors mean. In S. Sacks (Ed.), *On Metaphor*. University of Chicago Press.

Grady, J. E. (1997). THEORIES ARE BUILDINGS revisited. *Cognitive Linguistics, 8*(4), pp.267–290.

Hobbs, J. R. (1990). *Literature and cognition*. CSLI Lecture Notes, No. 21, Center for the Study of Language and Information, Stanford University.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought*, 2nd edition. Cambridge University Press.

Lakoff, G. & Turner, M. (1989). *More than cool reason: a field guide to poetic metaphor*. U. Chicago Press.

Martin, J. H. (1990). *A computational model of metaphor interpretation*. Academic Press.

Turner, M. & Fauconnier, G. (1995). Conceptual integration and formal expression. *Metaphor and Symbolic Activity, 10* (3), pp.183–204.

Veale, T. & Keane, M. T. (1997). The competence of sub-optimal structure mapping on 'hard' analogies. In *Procs. Int. Joint Conf. On Artificial Intelligence* (Nagoya, Japan), August 1997.

ALIGNMENT AND ABSTRACTION IN METAPHOR

Brian F. Bowdle

Department of Psychology
Indiana University
Bloomington, IN 47405
bbowdle@indiana.edu

INTRODUCTION

Metaphors establish mappings between concepts from disparate domains of knowledge. For example, in the metaphor *The mind is a computer*, an abstract entity is described in terms of a complex electronic device. It is widely believed that metaphors are a major source of knowledge change, and a great deal of research has examined how metaphors can enrich and illuminate concepts that would otherwise remain vague or ambiguous. However, there have been far fewer attempts to explain a second generative function of metaphors – namely, lexical extension. In this paper, I will discuss (1) how metaphoric mappings create new word meanings, and (2) how these new meanings are applied in subsequent metaphor processing. Before turning to these issues, however, it is necessary to consider the nature of metaphoric mappings in greater depth.

METAPHOR AND ANALOGY

Metaphors are traditionally viewed as comparisons between the target (a-term) and the base (b-term). According to several recent versions of this view, metaphors act to set up correspondences between isomorphic conceptual structures (e.g., Carbonell, 1981; Gentner, 1983; Gentner, Falkenhainer, & Skorstad, 1988; Indurkhya, 1987; Verbrugge & McCarrell, 1977). In other words, metaphor can be seen as a species of analogy.

Gentner's (1983) *structure-mapping theory* is among the most clearly articulated and extensively studied of these approaches to metaphor comprehension. Structure-mapping theory assumes that the act of comparison involves two stages: alignment and projection. The alignment process operates to create a maximal structurally consistent match between two representations that observes *one-to-one mapping* and *parallel connectivity* (Falkenhainer, Forbus, & Gentner, 1989). That is, each element of one representation can be placed in correspondence with at most one element of the other representation, and arguments of aligned relations are themselves aligned. A final constraint on the alignment process is *systematicity*: Alignments that form deeply interconnected structures, in which higher-order relations constrain lower-order relations, are preferred over less systematic sets of commonalities. Once a structurally consistent match between the target and base domains has been found, further predicates from the base that are connected to the common system can be projected to the target as *candidate inferences*.

To illustrate these processes, consider the metaphor *Men are wolves*. Given the simple target and base representations shown in Figure 1, structure-mapping theory predicts the following sequence of events in interpreting the metaphor. First, the relation *prey on*, which is shared by the target and base, is aligned. Next, the arguments of the relation are aligned by parallel connectivity: *wolves* à *men* and *animals* à *women*. Finally, predicates that are unique to the base but connected to the aligned structure (i.e., those predicates specifying that the predatory behavior is instinctive) are carried over to the target. Thus, the metaphor would be interpreted as meaning something like, "Men instinctively prey on women."

In many metaphors (as in analogies), the focus is on relational commonalities, and corresponding objects in the target and base need not be similar. Thus, in the above example, the alignment of the target *men* and the base *wolves* was determined primarily by the matching relation *prey on*. However, the way in which men prey on women is different from the way in which wolves prey on animals. This situation, in which matching predicates contain domain-specific differences, is typical of metaphors (e.g., Ortony, 1979; Tourangeau & Sternberg, 1981). Metaphoric mappings may therefore require *rerepresentation* in one or both terms. In particular, domain-specific features of matching predicates may be omitted so that the common structure is made more obvious (see Clement, Mawby, & Giles, 1994, for a review of this and other modes of rerepresentation).

METAPHOR AND POLYSEMY

Like analogies, metaphors lend additional structure to problematic target concepts, thereby making these concepts more coherent. However, this is not the only way in which metaphors can lead to knowledge change. Metaphors are also a primary source of polysemy – they allow words with specific meanings to take on additional, related meanings (e.g, Lakoff, 1987; Lehrer, 1990; Miller, 1979; Nunberg, 1979; Sweetser, 1990). For example, consider the word *roadblock*. There was presumably a time when this word referred only to a barricade set up in a road. With repeated metaphoric use, however, *roadblock* has acquired the secondary sense "anything that blocks progress" (as in *Fear is a roadblock to success*).

How do metaphors create new word meanings? One recent and influential proposal is that such lexical extensions are due to stable projections of conceptual structures and corresponding vocabulary items from one (typically concrete) domain of experience to another (typically abstract) domain of experience (e.g., Lakoff, 1987; Lehrer, 1990; Sweetser, 1990). On this view, the metaphoric meaning of a polyse-

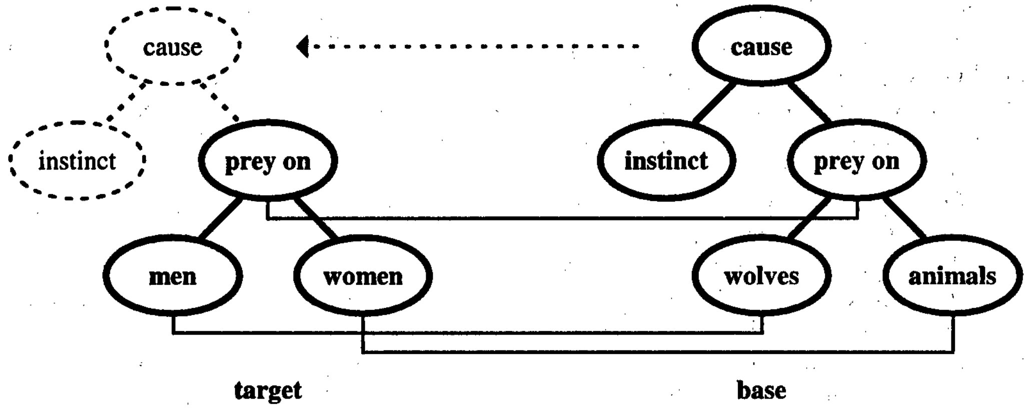


Figure 1. A structure-mapping interpretation of the metaphor Men are wolves.

mous word is understood directly in terms of its literal meaning.

I wish to consider an alternative account of the relationship between metaphor and polysemy – one that is based on the analogical approach to metaphor comprehension. The central idea is that structural alignment allows for the creation of abstract metaphoric categories, which may in turn be lexicalized as secondary senses of metaphor base terms (Bowdle, 1998; Bowdle & Gentner, 1995, in preparation; Gentner & Wolff, 1997).

The Induction of Metaphoric Categories

When a novel metaphor is first encountered, both the target and base terms refer to domain-specific concepts, and the metaphor is interpreted by (1) aligning the two representations, and (2) importing predicates from the base to the target, which then count as further matches. As a result of this comparison process, the common relational structure will increase in salience relative to domain-specific differences between the two representations. This highlighted system may in turn give rise to an abstract metaphoric category of which the target and base can be seen as instances. This is akin to the induction of domain-general problem schemas during the course of analogical problem solving (e.g., Gick & Holyoak, 1983; Novick & Holyoak, 1991; Ross & Kennedy, 1990).

On this view, metaphoric categories are created as a byproduct of the comparison process, and may be stored separately from the original target and base concepts. However, if a given metaphor base is repeatedly aligned with different targets so as to yield the same basic interpretation, the abstraction will become conventionally associated with the base term. At this point, the base term will be polysemous, having both a domain-specific meaning and a related domain-general meaning.

Of course, not just any metaphor can lead to lexical extension. Rather, the alignment of the target and base concepts must be able to suggest a coherent category. Mappings that focus on relational structures are therefore more likely to generate stable abstractions than mappings that focus on less systematic object descriptions (see also Ramscar & Pain, 1996; Shen, 1992). For example, the metaphor *The sun is a tangerine* elicits two common attributes of the target and base: Both are round, and both are orange in color. However, these two attributes are not systematically related. The metaphor is therefore unlikely to suggest a category of things that are round and orange in color, and it will not lead to lexical extension of the base term *tangerine*.

The Career of Metaphor

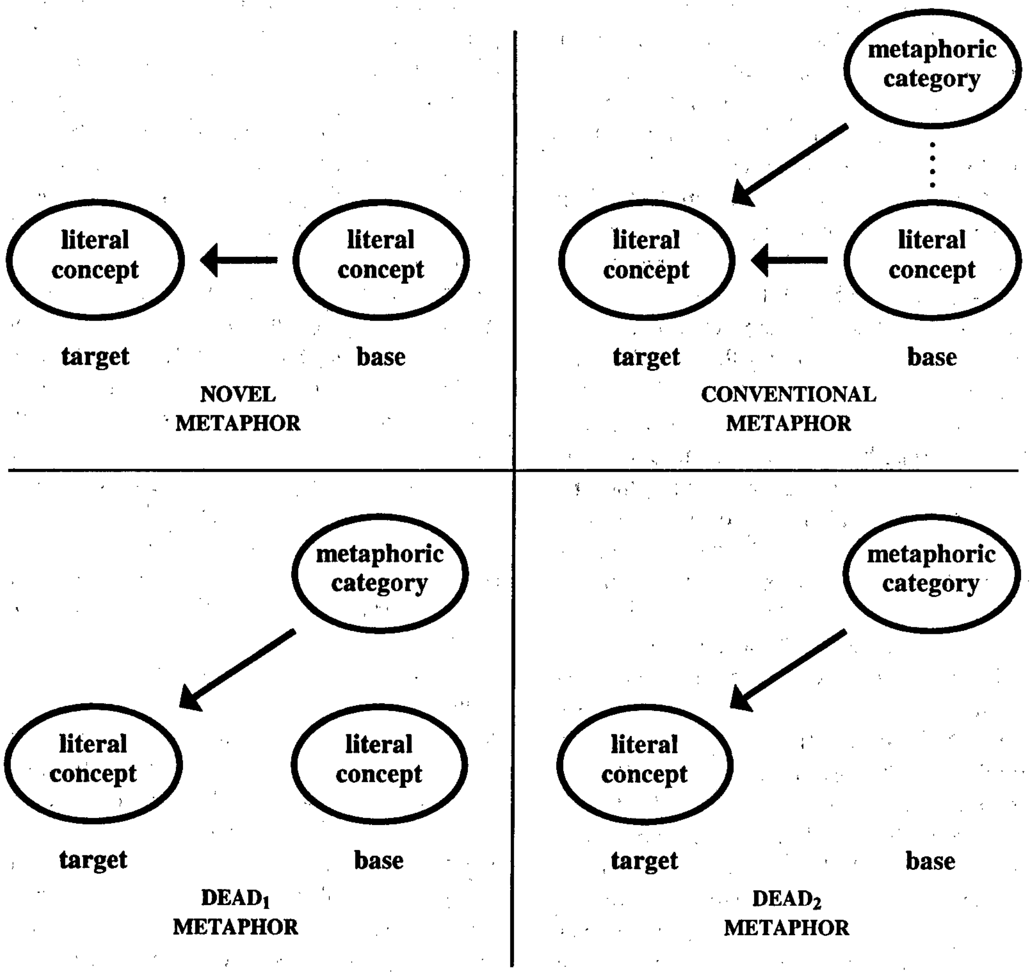
One of the key issues in metaphor research concerns how best to characterize differences between novel, conventional, and dead metaphors. The present account of the relationship between metaphor and polysemy suggests a representational distinction between these types of metaphors. The basic idea is that the conventionality of a metaphor is determined by (1) whether or not the base term evokes a metaphoric category, and (2) how this abstraction is related to the literal base concept.

The evolution from novel to dead metaphors is summarized in Figure 2. Novel metaphors involve base terms that refer to a domain-specific concept, but are not (yet) associated with a domain-general category. For example, the novel base term *glacier* (as in *Science is a glacier*) has a literal sense – "a large body of ice spreading outward over a land surface" – but no related metaphoric sense (e.g., "anything that progresses slowly but steadily").

In contrast, conventional metaphors involve base terms that refer both to a literal concept and to an associated metaphoric category. For example, the conventional base term *blueprint* (as in *A gene is a blueprint*) has two closely related senses: "a blue and white photographic print in showing an architect's plan" and "anything that provides a plan." Conventional base terms are polysemous, and the literal and metaphoric meanings are semantically linked due to their obvious similarity.

The ultimate conclusion of the career of metaphor occurs when the relationship between the derived metaphoric category and the original base concept is no longer recognized. At this stage, any expression using the metaphoric sense of the base term is a dead metaphor, and will not seem metaphoric. Figure 2 shows two possible types of dead metaphors. *Dead$_1$ metaphors* are similar to conventional metaphors, except that the two representations evoked by the base term are no longer semantically linked. That is, dead$_1$ base terms are homonymous rather than polysemous. For example, consider the statement *A university is a culture of knowledge*. Here, the word *culture* refers to a particular heritage or society, and its use seems quite literal. In fact, this sense of *culture* is a metaphoric extension of another commonly-known sense of the word: "a preparation for growth" (as in *the culture of the vine* or *bacteria culture*). However, these two meanings no longer seem related. This is perhaps because the once-abstract metaphoric category has, through repeated application to the domain of human affairs, acquired new domain-specific features.

Finally, *dead$_2$ metaphors* involve base terms that refer only to a derived metaphoric category – the original base concept no longer exists. An example of this is the dead$_2$ base term *blockbuster* (as in *The movie "Titanic" was a*



literal concept
literal concept
target
base
NOVEL METAPHOR
metaphoric category
literal concept
literal concept
target
base
CONVENTIONAL METAPHOR
metaphoric category
literal concept
literal concept
target
base
DEAD1 METAPHOR
metaphoric category
literal concept
target
base
DEAD2 METAPHOR

Figure 2. Four types of metaphors.

blockbuster), which means "anything that is highly effective or successful." However, most people are unaware that this word originally referred to a very large bomb that could demolish an entire city block.

PROCESSING IMPLICATIONS

Thus far, I have discussed how abstract metaphoric categories are created, and how these categories may be lexicalized as secondary senses of metaphor base terms. Not only does this account offer a means of distinguishing between novel, conventional, and dead metaphors, but it also has clear implications for the effects of conventionality on metaphor processing.

Consider again the career of metaphor summarized in Figure 2. In novel metaphors, both the target and base terms refer to domain-specific concepts at roughly the same level of abstraction. Novel metaphors will therefore be interpreted as comparisons, in which the target is structurally aligned with the base. In conventional metaphors, however, the base term is polysemous – it refers both to a domain-specific concept and to a related domain-general category. Conventional metaphors may therefore be interpreted either as comparisons, by aligning the target concept with the literal base concept, or as categorizations, by aligning the target concept with the metaphoric category named by the base term. Finally, in dead metaphors, only the metaphoric category named by the base will be applied to the target – the original base concept either seems irrelevant (dead$_1$ metaphors) or is no longer available (dead$_2$ metaphors).

Thus, as metaphors become increasingly conventional, there is a shift in mode of processing from comparison to categorization (Bowdle, 1998; Bowdle & Gentner, 1995, in preparation; Gentner & Wolff, 1997). This is consistent with a number of recent proposals, according to which the interpretation of novel metaphors involves sense creation, but the interpretation of conventional metaphors involves sense retrieval (e.g., Blank, 1988; Blasko & Connine, 1993; Giora, 1997; Turner & Katz, 1997). On the present view, the senses retrieved during conventional metaphor comprehension are abstract metaphoric categories.

Experimental Evidence

To gain direct evidence for the processing shift predicted by the career of metaphor, Dedre Gentner and I have recently conducted a series of experiments comparing the comprehension and evaluation of novel and conventional figurative statements (Bowdle & Gentner, 1995, in preparation). Central to the logic of these experiments was the distinction between metaphors and similes.

Nominal metaphors (figurative statements of the form *X is Y*) can often be paraphrased as similes (figurative statements of the form *X is like Y*). For example, one can say both *The mind is a computer* and *The mind is like a computer.* This linguistic alternation is interesting because metaphors are grammatically identical to literal categorization statements (e.g., *A sparrow is a bird*), and similes are grammatically identical to literal comparison statements (e.g., *A sparrow is like a robin*). Assuming that form typically follows function in both literal and figurative language, metaphors and similes may tend to promote different comprehension strategies. Specifically, metaphors should invite classifying the target as a member of a category named by the base, whereas similes should invite comparing the target to the base. This makes the metaphor-simile distinction a valuable tool for examining the use of comparison and categorization during figurative language comprehension.

Grammatical Form Preferences. If conventionalization results in a processing shift from comparison to categorization, then there should be a corresponding shift at the linguistic level from the comparison (simile) form to the categorization (metaphor) form. We gave subjects novel and conventional figurative statements in both grammatical forms, and asked which form they preferred for each statement. Subjects were also given statements in which the target was literally similar to the base (e.g., *lemon* à *orange*) – for which the comparison form is most natural

– and statements in which the target was a member of a literal category named by the base (e.g., *whale* à *mammal*) – for which the categorization form is most natural.

As predicted, subjects preferred similes more strongly for novel than for conventional figurative statements. Indeed, the preference for the comparison form was as great for novel figuratives as for statements in which the target and base were literally similar. However, subjects showed no strong preference for expressing conventional figurative statements as similes or as metaphors. This is consistent with the claim that, because conventional base term refer both to a literal concept and to a related metaphoric category, conventional figuratives may be interpreted either as comparisons or as categorizations.

Comprehension Times. The career of metaphor also makes clear predictions about the online comprehension of novel and conventional figurative statements. One prediction is that, if conventionalization results in a processing shift from comparison to categorization, then conventional figuratives should be easier to interpret than novel figuratives. Because metaphoric categories will be informationally sparser than the literal concepts they were derived from, mappings between a target and a metaphoric category will be computationally less costly than mappings between a target and a literal base concept. In fact, previous studies have confirmed that conventional metaphors are comprehended more rapidly than novel metaphors (e.g., Blank, 1988; Blasko & Connine, 1993).

A second and more interesting prediction concerns the effects of conventionality on the relative comprehension times of metaphors and similes. If novel figurative statements are interpreted strictly as comparisons, then novel similes should be easier to comprehend than novel metaphors. This is because only the simile form directly invites comparison. At the same time, if conventional figurative statements can be interpreted either as comparisons or as categorizations, then conventional metaphors should be easier to comprehend than conventional similes. The metaphor form invites categorization, and will therefore promote a relatively simple alignment between the target and the abstract metaphoric category named by the base. The simile form invites comparison, and will therefore promote a more complex alignment between the target and the literal base concept.

We collected subjects' comprehension times for novel and conventional figurative statements phrased either as metaphors or as similes. The results were as predicted by the career of metaphor. First, conventional figurative statements were interpreted faster than novel figurative statements. Second, there was an interaction between conventionality and grammatical form: Novel similes were faster than novel metaphors, but conventional metaphors were faster than conventional similes.

Metaphoricity Ratings. What makes some mappings seem metaphoric and other mappings seem literal? One possibility is that metaphoricity is due to rerepresentation, in which distinct domain-specific features of matching predicates are omitted so that the common structure is made more obvious (Bowdle, 1998). This is consistent with the observation that metaphors and similes typically involve mappings between concepts from different ontological domains, whereas literal comparisons and literal categorizations typically involve mappings between concepts from the same ontological domain.

This view of the relationship between rerepresentation and metaphoricity suggests a further test of the career of metaphor. If conventionalization results in a processing shift from comparison to categorization, then novel figurative statements should seem more metaphoric than conventional figurative statements. Because the predicates of literal base concepts will be more domain-specific than those of abstract metaphoric categories, they will require more rerepresentation when matched with domain-specific predicates in a target concept.

A further prediction concerns how conventionality affects the relative metaphoricity of metaphors and similes. If both novel metaphors and novel similes are interpreted by aligning the target with the same literal base concept, then both grammatical forms should seem equally metaphoric. At the same time, if conventional metaphors promote aligning the target with an abstract metaphoric category, but conventional similes promote aligning the target with a literal base concept, then conventional similes should seem more metaphoric than conventional metaphors – the simile will initiate a mapping that requires a greater degree of rerepresentation. Note that this prediction is contrary to the traditional (and previously untested) assumption that metaphors are more metaphoric than similes.

We gave subjects novel and conventional figurative statements phrased either as metaphors or as similes, and asked them to rate the metaphoricity of each statement. The results were as predicted by the career of metaphor. First, novel figurative statements were rated as more metaphoric than conventional figurative statements. Second, there was an interaction between conventionality and grammatical form: Novel metaphors and similes were equally metaphoric, but conventional similes were more metaphoric than conventional metaphors.

CATEGORIZATION MODELS OF METAPHOR

One of the central claims made in this paper is that as metaphors are conventionalized – that is, as they increasingly rely on the application of stable abstractions – there is a shift in mode of processing from comparison to categorization. However, several theorists have recently argued that all metaphors are essentially categorizations (e.g., Glucksberg & Keysar, 1990; Glucksberg, McGlone, & Manfredi, 1997; Honeck, Kibler, & Firment, 1987; Kennedy, 1990). On this view, the original target and base concepts of a novel metaphor are never directly aligned. Rather, the metaphor is interpreted by (1) deriving an abstract metaphoric category from the base concept alone, and (2) applying this category to the target.

The experimental evidence summarized above casts doubt on these processing claims. Novel metaphors appear to be interpreted strictly as comparisons, in which the target is structurally aligned with the base. Although novel metaphoric mappings may create abstract metaphoric categories, these categories initially arise as a byproduct of the comparison process. Only when a metaphoric category has become conventionally associated with the base term of a metaphor can the statement be interpreted as a categorization. Assuming that the metaphor is not dead, however, it may still be interpreted as a comparison between the target and the original base concept.

CONCLUSIONS

By viewing metaphors as analogies, two generative functions of metaphors can be explained – namely, the structural enhancement of target concepts, and the lexical extension of base terms. In this paper, I have focused on the latter of these two functions, and have discussed the relationship between polysemy and conventionality in metaphors. The career of metaphor outlined here offers a unified approach to metaphor processing.

REFERENCES

Blank, G. D. (1988). Metaphors in the lexicon. *Metaphor and Symbolic Activity*, 3, 21-36.

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 295-308.

Bowdle, B. F. (1998). *Conventionality, polysemy, and metaphor comprehension.* Unpublished doctoral dissertation, Northwestern Universiy.

Bowdle, B. F., & Gentner, D. (1995). *The career of metaphor.* Poster given at the Thirty-Sixth Annual Meeting of the Psychonomic Society, Los Angeles, CA.

Bowdle, B. F., & Gentner, D. (in preparation). The career of metaphor.

Carbonell, J. G. (1981). Invariance hierarchies in metaphor interpretation. In *Proceed-*

ings of the Third Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Clement, C. A., Mawby, R., & Giles, D. E. (1994). The effects of manifest relational similarity on analog retrieval. *Journal of Memory and Language*, 33, 396-420.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41, 1-63.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.

Gentner, D., Falkenhainer, B., & Skorstad, J. (1988). Viewing metaphor as analogy. In D. H. Helman (Ed.), *Analogical reasoning*. New York: Kluwer.

Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, 37, 331-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 8. 183-206.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, 97, 3-18.

Glucksberg, S., McGlone, M. S., & Manfredi, D. (1997). Property attribution in metaphor comprehension. *Journal of Memory and Language*, 36, 50-67.

Honeck, R. P., Kibler, C. T., & Firment, M. J. (1987). Figurative language and psychological views of categorization: Two ships in the night? In R. E. Haskell (Ed.), *Cognition and symbolic structures*. Norwood, NJ: Ablex.

Indurkhya, B. (1987). Approximate semantic transference: A computational theory of metaphor and analogy. *Cognitive Science*, 11, 445-480.

Kennedy, J. M. (1990). Metaphor – its intellectual basis. *Metaphor and Symbolic Activity*, 5, 115-123.

Lakoff, G. (1987). *Women, fire, and dangerous things.* Chicago: University of Chicago.

Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics*, 1, 207-246.

Miller, G. A. (1979). Images and models, similes and metaphors. In A. Ortony (Ed.), *Metaphor and thought.* New York: Cambridge University.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 398-415.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, 3, 143-184.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, 86, 161-180.

Ramscar, M. J. A., & Pain, H. G. (1996). Can a real distinction be made between cognitive theories of analogy and categorization? In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 42-55.

Shen, Y. (1992). Metaphors and categories. *Poetics Today*, 13, 771-794.

Sweetser, E. (1990). *From etymology to pragmatics.* New York: Cambridge University.

Tournangeau, R., & Sternberg, R. J. (1981). Aptness in metaphor. *Cognitive Psychology*, 13, 27-55.

Turner, N. E., & Katz, A. N. (1997). The availability of conventional and of literal meaning during the comprehension of proverbs. *Pragmatics and Cognition*, 5, 199-233.

Verbrugge, R. R., & McCarrell, N. S. (1977). Metaphoric comprehension: Studies in reminding and resembling. *Cognitive Psychology*, 9, 494-533.

Evidence for Metaphoric Representation: Understanding Time

Lera Boroditsky
Stanford University
Psychology, Bldg. 420
Stanford, CA 94305
lera@psych.stanford.edu

ABSTRACT

These experiments evaluated the claim that abstract conceptual domains are organized and structured on-line as metaphorical mappings from conceptual domains grounded directly in experience. One hypothesis is that the conceptual domain of time is systematically organized in terms of the more concrete and familiar domain of space. I focus on relational similarities between the conceptual domains of space and time, consider a number of explanations of how these similarities may have come about, and describe a set of experiments designed to distinguish between these explanations. The results indicated that people indeed use spatial schemas on-line to understand and organize the conceptual domain of time. These results provide some of the first empirical evidence for metaphoric representation.

INTRODUCTION

One of the burdens of providing a good theory of mental representation is to explain how a representational store as heterogeneous, sophisticated, and abstract as the human concepticon could possibly emerge from physical experience with the world. One solution proposed by Lakoff and Johnson (1980) argues that our conceptual system is structured around a small set of experiential concepts (concepts that emerge directly out of experience and are defined in their own terms). These fundamental experiential concepts include a set of basic spatial relations (e.g. up/down, front/back), a set of physical ontological concepts (e.g. entity, container), and a set of basic experiences or actions (e.g. eating, moving). According to Lakoff, all other concepts that do not emerge directly out of physical experience must be metaphoric in nature. Lakoff and colleagues further propose that these metaphoric, or abstract concepts are understood and structured through on-line metaphorical mappings from a small set of fundamental experiential concepts. This paper aims to test the psychological validity of the metaphoric theory of mental representation.

Lakoff and his colleagues have noted the presence of many large-scale systems of conventional conceptual metaphors; cases in which language from one domain is used in other domains (Lakoff & Johnson, 1980, Lakoff & Kovecses, 1987). These conventional metaphors can often be characterized as belonging to a particular source-to-target mapping: e.g., MIND IS A CONTAINER, IDEAS ARE FOOD. In keeping with the IDEAS ARE FOOD schema, for example, a reader might be reluctant to "swallow Lakoff's claim" because they haven't yet gotten to "the meaty part of the paper," or because they "just can't wait to really sink their teeth into the theory."

Such linguistic patterns suggest that many conceptual domains can be described systematically in terms of more tangible and familiar domains (as in the IDEAS ARE FOOD schema described above). However, whether these large-scale schemas are psychologically real conceptual systems or post-hoc theoretical constructs remains an open question.

In this paper, I will highlight a set of relational similarities between the conceptual domains of space and time, consider several explanations of how these similarities may have come about, and describe two experiments that distinguish between these explanations. The described experiments will directly test the psychological validity of Lakoff's claim that abstract conceptual domains are structured by metaphorical mappings from more concrete experiential domains. Let us now focus on the domain of time.

SPATIAL METAPHORS FOR TIME

We often talk about time in terms of space. Whether we are looking *forward* to a brighter tomorrow, proposing theories *ahead* of our time, or falling *behind* schedule, we are relying on terms from the domain of space to talk about time. There is an orderly and systematic correspondence between the domains of time and space in language (Bennett, 1975; Clark, 1973; Lehrer, 1990; Traugott, 1978).

The correspondences between space and time in language might give us insight into how we mentally represent time. Let us focus on the event-sequencing aspect of conceptual time, the system whereby events are temporally ordered with respect to each other and to the speaker (e.g. "The worst is behind us" or "Thursday is before Saturday.") In order to capture the sequential order of events, time is generally conceived as a one-dimensional, directional entity. The spatial terms we import to talk about time are also one-dimensional, directional terms such as *ahead/behind*, *up/down*, as opposed to multi-dimensional or symmetric terms such as *shallow/deep, left/right.* This pattern is stable across languages, and overall, spatial terms referring to front/back relations are the ones most widely borrowed into the domain of *time* cross-linguistically (Clark, 1973; Traugott, 1978).

As most abstract domains, the domain of time can be described through more than one metaphor. In English, there are two dominant *space —>time* metaphoric systems (Clark, 1973; Lakoff & Johnson, 1980). The first system can be termed the *ego-moving* metaphor, where "ego" or the observer's context progresses along the time-line toward the future as in *"We are coming up on Christmas"* (see Figure 1a). The second system is the *time-moving* metaphor. In this metaphor, a time-line is conceived of as a river or conveyor belt on which events are moving from the future to the past as in *"Christmas is coming"* (see Figure 1b). These two systems lead to different assignments of front/back to a time-line (Clark, 1973; Fillmore, 1979; Lakoff & Johnson, 1980; Traugott, 1978). In the ego-moving system, front is assigned to the future or later event (e.g. "The war is *behind* us" or "His whole future is *before* him"). In the time-moving system, front is assigned to a past or earlier event (e.g. "I will see you *before* 4 o'clock" or "The reception *after* the talk.")

Although the apparent systematicity and coherence of the ego-moving and time-moving systems in temporal language is compelling, a priori it is not clear that any structured conceptual schemas are necessary to process metaphoric expressions about time.

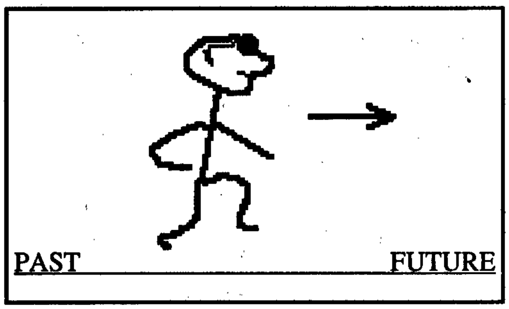


Figure 1a. Ego-moving schema

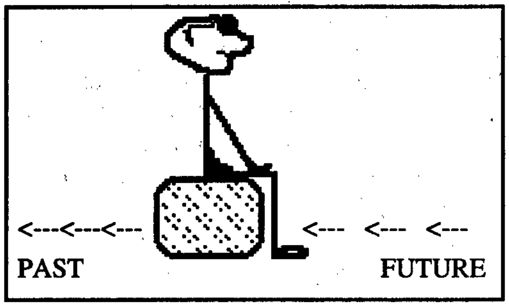


Figure 1b. Time-moving schema

It could be the case that "metaphoric" expressions are simply polysemous expressions. That is, the "metaphoric" meaning is stored as a secondary meaning in the lexical entry of the base term. A word like "ahead," for example, might have two (or more) word senses associated with it: 'in front of spatially' and 'in front of temporally'. If this is the case, one need not carry out any structured mapping between domains in order to understand what "ahead" means in a temporal context.

There is some evidence for this alternative hypothesis. For example, Glucksberg, Brown, and McGlone (1993) showed that people do not access the "anger is heat" metaphor when processing conventional idioms such as "lose one's cool."

It is possible, then, that the ego-moving and time-moving metaphors are only language-deep — etymological relics, not psychologically real conceptual schemas. In order to establish the psychological reality of these event-sequencing schemas, we must first be able to empirically distinguish between expressions that are simply polysemous, and those that are processed as parts of globally consistent conceptual schemas.

EVIDENCE FOR TWO DISTINCT EVENT-SEQUENCING SCHEMAS

To investigate whether the ego-moving and time-moving conceptual schemas are used in real-time language comprehension, Gentner, Imai, and Boroditsky (in preparation) measured processing time for statements using event-sequencing expressions presented either consistently or inconsistently with respect to either the ego-moving or the time-moving schema. They reasoned that if temporal expressions were processed as parts of globally consistent conceptual schemas, then processing should be fluent if the expressions are kept consistent to one schema (processing time should remain constant). If the schemas are switched, however, processing should be disrupted, and processing time should increase as it would take extra time to discard the old conceptual structure and set up a new one.

Participants were presented with a block of temporal statements that were either consistent to one schema, or switched between the ego-moving and time-moving schemas. For each statement (e.g. Christmas is six days before New Year's Day), participants were given a time-line of events (e.g. Past.........New Year's Day.........Future), and had to place an event (in this case Christmas) on the timeline. Response time data showed that switching schemas did indeed increase processing time.

Another study was conducted at Chicago's O'Hare airport where participants were passengers not aware of being in a psychological study. Participants were approached by the experimenter and asked a priming question in either the ego-moving form (Is Boston ahead or behind us in time?) or the time-moving form (Is it earlier or later in Boston than it is here?). After the participant answered, the experimenter asked the target question (So should I turn my watch forward or back?) which was consistent with the ego-moving form. Response times for the target question were collected with a stopwatch disguised as a wristwatch. Once again, response times for consistently primed questions were shorter than for inconsistently primed questions. Switching schemas increased processing time. These results suggest that there are two distinct conceptual schemas that are involved in sequencing events in time.

Converging evidence comes from a study by McGlone, Harding, and Glucksberg (1994). Participants answered blocks of questions about days of the week phrased in either the ego-moving or the time-moving metaphor. The ego-moving blocks were composed of statements like "We passed the deadline yesterday." The time-moving blocks were composed of statements like "The deadline was passed yesterday." For each statement participants were asked to indicate the day of the week that the event in the statement had occurred or will occur. At the end of each block, participants were presented with an ambiguous temporal statement such as "Friday's game has been moved forward a day," and were asked to perform the same task. The above statement is ambiguous because it could be interpreted using one or the other schema to yield different answers. McGlone et al. found that participants in the ego-moving condition tended to disambiguate the above statement in an ego-moving-consistent manner (thought the game was on Saturday), and participants in the time-moving condition tended to disambiguate in a time-moving-consistent manner (thought the game was on Thursday).

These studies provide evidence for the existence of two distinct, globally consistent conceptual schemas for sequencing events in time. The challenge now is to show that these two large-scale schemas are imported from the experiential domain of space to the abstract domain of time during real-time processing, as the metaphorical representation hypothesis would imply.

A reasonable alternative to the metaphorical representation hypothesis was proposed by Murphy (1996) who argued that all domains are represented directly, not metaphorically. According to Murphy's Structural Similarity hypothesis, metaphorical language arises from pre-existing structural similarities between two domains. The two domains are represented separately, but are quite similar, and it is this conceptual similarity that allows people to construct understandable verbal metaphors.

Before we can empirically distinguish between these two hypotheses, we need to make explicit the analogy between the schemas used to order events in time, and the schemas used to order objects in space. That is, if some structures in time are metaphors from space, what are the spatial schemas that serve as the base of this metaphor?

STRUCTURAL SIMILARITIES BETWEEN SPACE AND TIME

Many structural similarities exist between the conceptual domains of space and time. A set of spatial analogs for the ego-moving and time-moving schemas is proposed below.

The Ego-moving schema

According to the ego-moving schema, events in the domain of time are ordered with respect to the observer's direction of motion. The front of an ego-moving scenario is assigned as the furthest point in the observer's direction of motion. Since in the domain of time the observer is inevitably moving from the past to the future, front is assigned to future or later events. An analogous schema exists for ordering objects in a line (see Figure 2). When an observer moves along a path, objects are ordered according to the direction of motion of the observer. In the example in Figure 2, the dark can is said to be in front because it is further along in the observer's direction of motion.

Time- or Object-moving

According to the time-moving schema, events in time are ordered based on the direction of motion of time. The front of a time-moving scenario is the furthest point in the direction of motion of time. Since time inevitably moves from the future to the past, front is assigned to past or earlier events. Once again an analogous system exists for ordering objects in space (see Figure 3). When two objects (without intrinsic fronts) are moving, they are assigned fronts based on their direction of motion. The front here, just as in the domain of time, is assigned to the leading part of an object in the direction of motion. The light-colored widget is said to be in front because it is further along in the widgets' direction of motion.

We are now in a position to ask whether the same relational schemas are used to sequence objects in space and events in time. If the same schemas are indeed used by both domains, then we should be able to differentially prime particular spatial schemas to affect how people think about time.

In the first experiment we were interested in whether making subjects think about spatial relations in a particular way, would affect how they think about time. We primed either the ego-moving or the object-moving spatial schemas by asking subjects to answer some questions about the spatial relations of objects in a picture. We then asked subjects to interpret an ambiguous temporal statement such as "Next Wednesday's meeting has been moved forward two days." If the above sentence is interpreted using the ego-moving schema, then *forward* is in the direction of motion of the observer, and the meeting should now fall on a Friday. In the time-moving interpretation, however, *forward* is in the direction of motion of time, and the meeting should now be on a Monday.

If the domains of time and space do indeed use the same relational schemas, subjects primed in the ego-moving spatial frame of reference should prefer the ego-moving perspective for reasoning about events in time, and should think that the meeting will be on Friday. Subjects primed in the object-moving frame of reference should prefer the time-moving interpretation and think that the meeting will be on Monday. However, if the domains of space and time are represented separately, then spatial primes should have no effect on the way subjects think about time.

EXPERIMENT 1

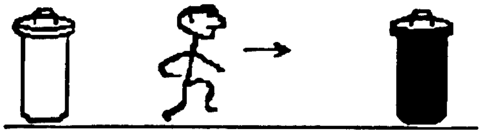
METHOD

Participants

63 Stanford University undergraduates participated in this study as part of a course requirement.

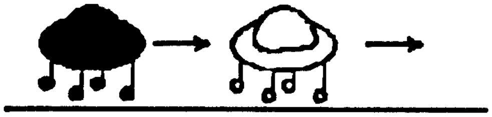
Materials & Design

A two-page questionnaire was constructed. The first page contained four TRUE/FALSE priming questions about spatial scenarios. Scenarios used either the ego-moving frame of reference (see Figure 2), or the object-moving frame of reference (see Figure 3). These two frames of reference were predicted to map onto the ego-moving and time-moving perspectives in time respectively. On a separate page, participants were asked to read an ambiguous temporal sentence (e.g. "Next Wednesday's meeting has been moved forward two days.") and report on which day the meeting has been rescheduled. A control group of subjects responded to the target sentence without having seen a prime. All subjects also provided a confidence score for their answer to the target question on a scale of 1 to 5 (1=not at all confident, 5=very confident).



The dark can is in front of me.

Figure 2. Sample ego-moving scenario.



The light widget is in front of the dark widget.

Figure 3. Sample object-moving scenario.

Procedure

Participants completed the two-page questionnaire individually with no time restrictions. The two pages of the questionnaire were imbedded in a large questionnaire packet containing questions unrelated to this study. No overt connection was made between the two pages of the questionnaire pertaining to this study.

Results

As predicted by the metaphorical representation hypothesis, participants responded in a prime-consistent manner. Of the participants primed in the ego-moving frame of reference, 73.3% thought that the meeting was on Friday, and 26.7% thought it was on Monday. The reverse pattern was true for the participants primed in the object-moving frame of reference. Only 30.8% of the participants primed in the object-moving frame of reference thought the meeting was on Friday, and 69.2% thought it was on Monday. A Chi-squared statistic confirmed the effect of consistency (Chi = 5.2, p<.05). Control participants (who had not seen any primes) were evenly split between Monday (45.7%) and Friday (54.3%).

An additional measure of confidence confirmed the large consistency effect. A confidence score was computed for each subject by scoring a prime-consistent response as a +1, and a prime-inconsistent response as a -1 and multiplying by the confidence rating that had been provided by the subject on a scale of 1 to 5. Under the null hypothesis, the mean confidence score should equal 0. The mean observed confidence score for the primed conditions was 2.14, significantly higher than 0 (t=2.81, p<.01), once again confirming the consistency effect. For the unprimed or control condition, the mean confidence score did not differ from the null prediction (Mean = -0.23).

There was, however, one concern about the spatial scenario pictures used in Experiment 1. Besides, the difference in underlying structure, there were several superficial differences between the ego-moving and the object-moving pictures used. The ego-moving pictures always contained three entities, always had a person in them, and contained only one arrow indicating direction of motion. The object-moving pictures, on the other hand, only contained two entities, never contained a person, and always had two arrows indicating direction of motion. It is possible that any of these superficial differences could affect they way people responded to the target question. We conducted a follow experiment to address the above issues.

EXPERIMENT 1A

In the follow-up experiment, we redesigned the stimuli to minimize these superficial differences between the ego-moving and object-moving scenarios (see Figures 4-5). A different group of 71 Stanford undergraduates participated in the follow-up experiment. Just as in Experiment 1, there was a significant effect of schema consistency (Chi = 6.28, p<.05). Subjects primed with ego-moving pictures, chose the ego-moving response (Friday) 63% of the time. Subjects primed with the object-moving pictures chose the time-moving response (Monday) 67% of the time.

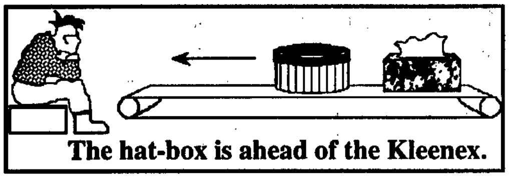

The hat-box is ahead of the Kleenex.

Figure 4. Sample object-moving scenario for Exp. 1a

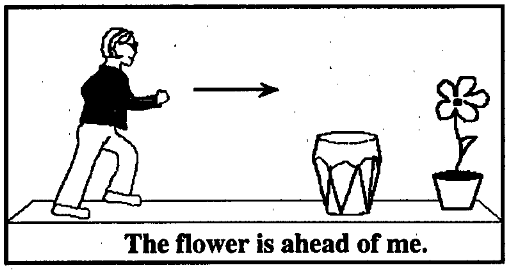

The flower is ahead of me.

Figure 5. Sample ego-moving scenario for Exp. 1a

In both studies, subjects were influenced by the primed spatial schemas when trying to solve a problem about time. These findings strongly suggest that structured relational information is shared between the domains of space and time.

Discussion

In Experiment 1 and the follow-up experiment we found that priming particular spatial schemas can affect how people think about time. Participants chose to disambiguate a sentence about time in a manner that was consistent with a recently used spatial schema. With these finding in hand, we can reject Murphy's Structural Similarity hypothesis that states that the domains of space and time, though similar, are not related. Experiments 1 and 1a provide strong evidence for the claim that the domains of space and time share relational structure. However, it is too early to conclude that time is understood and structured as a metaphor from space. There are two concerns.

First, since our data came from a questionnaire, we have no direct measurements of the real-time processing that went on while subjects were answering our questions. If we are to claim that spatio-temporal expressions are processed on-line as metaphorical mappings, we must be able to demonstrate that schema consistency has some effect on real-time processing. To address this concern, we need to design a more controlled laboratory task that will allow us to assess subjects' on-line processing.

The second concern has to do with how the ego-moving and time-moving schemas are represented. So far we have established that the domains of space and time are conceptually related, and that they share some relational schemas. These findings are consistent with the metaphorical representation hypothesis that structured relational information is stored in the domain of space and mapped to the domain of time. However, an alternative explanation of our results is that there are generic, domain-independent schemas that are shared by both domains. If we are to claim that abstract domains like time are understood as metaphors from concrete experiential domains like space, then we must be able to show that there is directional transfer between the two domains. We must show that information is transferred from space to time, and not from time to space. Under the metaphorical representation hypothesis, the schema-consistency effect should be asymmetric; there should be a greater effect of schema-consistency when the transfer is from space to time, than from time to space. Experiment 2 was designed to address the above two concerns.

In Experiment 2, subjects' response times to questions about spatial and temporal relations were measured. Each target question was preceded by two prime questions that used either the same relational schema as the target (a Consistent trial) or used a different relational schema (an Inconsistent trial). We also varied the domains from which the target and prime questions were drawn so that sometimes spatial primes were followed by target questions about time, and other times temporal primes were followed by target questions about space. We also included trials where spatial primes were followed by spatial targets, and temporal primes were followed by temporal targets. These trials were necessary as manipulation checks; we must be able to demonstrate that our stimuli can produce an effect of consistency within a domain before we can interpret consistency effects across domains.

EXPERIMENT 2

Predictions

Under the metaphorical representation hypothesis, we would predict that subjects should be slower to respond to inconsistently primed items when temporal targets are preceded by spatial primes. However, there should be no effect of consistency when spatial targets are preceded by temporal primes. The exact predictions are below:

Spatial primes to spatial targets. When schema-inconsistent primes are used, the primed inconsistent schema will interfere with processing and processing time will increase.

Spatial primes to temporal targets. When schema-inconsistent spatial primes are used, the inconsistent spatial schema will become very available. Since spatial schemas are used on-line for understanding temporal scenarios, this inconsistent schema will disrupt processing causing processing time to increase.

Temporal primes to temporal targets. When schema-inconsistent temporal primes are used, the product of the mapping that was necessary to process the prime scenarios will become most available. The product of this mapping will be the inconsistent set of correspondences between the domains of space and time. This inconsistent set of correspondences will interfere with processing, causing processing time to increase.

Temporal primes to spatial targets. When schema-inconsistent temporal primes are used, what will become most available is the product of the mapping that was necessary to process the prime scenarios. The product of this mapping will be the inconsistent set of correspondences between the domains of space and time. This set of correspondences should have no effect on the processing of a spatial scenario, since the domain of space is represented directly, and does not depend on the domain of time.

METHODS

Participants

34 Stanford University undergraduates participated in this study in order to fulfill a course requirement.

Materials

The experiment used 128 prime questions and 32 target questions. All questions were TRUE/FALSE. Each prime question appeared only once. Each target question appeared twice, once primed Consistently, and once primed Inconsistently.

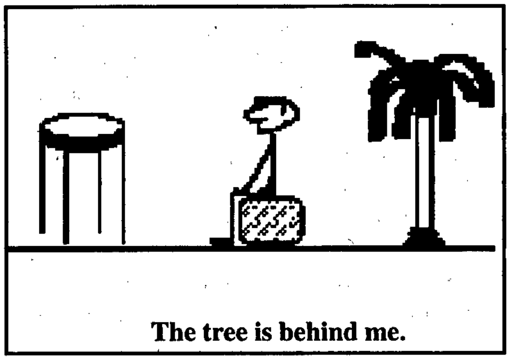
Time Questions: 64 statements about months of the year were constructed to use as primes. Half of these statements were phrased using the ego-moving schema (e.g. "In March, May is ahead of us."), and the other half used the time-moving schema (e.g. "March comes before May.") Also, half of the statements were TRUE and half were FALSE. Half of the statements referred to months that are "ahead" or "before", and half of the statements referred to months that are "behind" or "after". All of these variations were fully crossed into eight types of primes. This was done to insure that the task was difficult enough that subjects would not be able to develop a heuristic to answer the questions. In addition, 16 statements about months of the year were constructed to use as target questions. These statements were always TRUE, used either the ego-moving, or the time-moving schema, and always referred to months that are "ahead" or "before".

Space Questions: 64 spatial scenarios were constructed to use as primes. Each scenario consisted of a picture and a sentence. Half of these scenarios used the ego-moving schema, and the other half used the object-moving schema. Also, half of the sentences were TRUE descriptions of the spatial relations portrayed in the picture and half were FALSE descriptions. Half of the statements referred to objects that were "in front", and half referred to objects that are "behind". All of these variations were fully crossed into eight types of primes. Also, left/right orientation of the pictures was counterbalanced across these variations.

In addition, 16 spatial scenarios were constructed to use as target questions. Sentences in these scenarios were always TRUE descriptions of the picture, used either the ego-moving, or the object-moving schema, and always referred to objects that are "in front". Sample items can be found in Figure 6.

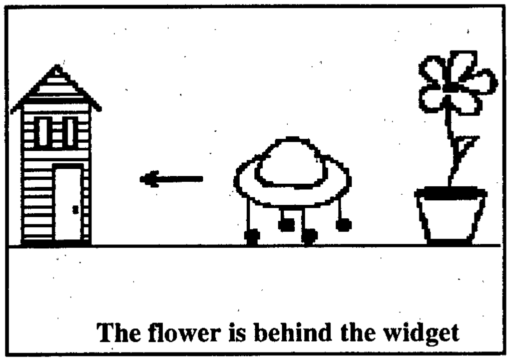
Design

Overall, the experiment has a three factor fully crossed within subjects design. The design is 4 (transfer type) X 2 (consistency) X 2 (target type). There were four levels of transfer type: (1) "space-to-space" - transfer



The tree is behind me.

Figure 6a. Sample ego-moving spatial scenario.



The flower is behind the widget

Figure 6b. Sample object-moving spatial scenario.

from spatial primes to spatial targets; (2) "space-to-time" - transfer from spatial primes to temporal targets; (3) "time-to-time" - transfer from temporal primes to temporal targets; and (4) "time-to-space" - transfer from temporal primes to spatial targets. There were two levels of consistency: (1) consistent - the primes and targets belong to the same schema; or (2) inconsistent - the primes and targets belong to different schemas. There were two levels of target type: (1) ego-moving; and (2) object/time-moving.

Each participant completed a short practice session and 64 experimental trials. Each trial was composed of two prime questions and one target question. Each target was presented twice, once in a Consistent trial, and once in an Inconsistent trial. A trial was Consistent when the prime questions and the target question belonged to the same schema (e.g. ego-moving prime, ego-moving target). A trial was Inconsistent when the prime questions and the target question belonged to different schemas (e.g. ego-moving prime, time-moving target). The critical measure was the effect of consistency on the response time to the same target question by the same subject. Trials were randomized for each subject with the constraint that the order of consistent and inconsistent presentations of the same target was counterbalanced across subjects. For each subject, consistent and inconsistent items appeared first and second equally often.

Procedure

Participants were tested individually. Participants completed a short practice session followed by 64 experimental trials. Upon a participant's completion of the practice session, the experimenter provided feedback, and repeated the instructions. There was no feedback for the 64 experimental trials.

In each trial, the participant saw 2 prime questions followed by one target question. Participants did not know that the experiment was broken up into such trials, nor could they figure it out just from being in the experiment. For each question a participant needed to make a TRUE/FALSE response. There was a response deadline of six seconds.

Results

Results are summarized in Figures 7-10. As predicted by the metaphorical representation hypothesis, subjects showed a schema consistency effect when the direction of transfer was from space to time, but not from time to space. Subjects also showed within-domain consistency effects in both the space-to-space transfer condition, and the time-to-time transfer condition. For each transfer type, a three-factor (2 Consistency X 2 Target type X 34 Subjects) GLM analysis was conducted. For each comparison there was a significant effect of subjects which is to be expected due to large individual differences in reaction time. Error responses were omitted from all analyses.

Within-domain schema consistency

Space-to-space: See Figure 7. Subjects responded significantly faster to Consistently presented targets (mean RT = 1590 msecs), than to Inconsistently presented targets (mean RT = 1700 msecs) when both the prime and target questions came from the domain of space (F= 5.01, p<.05). Establishing this within-domain consistency effect was necessary as a manipulation check. There was no interaction between Target type and Consistency. This means that both ego-moving and object-moving targets benefited equally from Consistency.

Time-to-time: See Figure 8. Subjects responded faster to Consistently presented targets (mean RT = 1761 msecs), than to Inconsistently presented targets (mean RT = 1896 msecs) when both the prime and target questions came from the domain of time (F=4.42, p<.05). Establishing this within-domain consistency effect was necessary as a manipulation check. There was no effect of Target type, and no interaction between Target type and Consistency. This means that both ego-moving and time-moving targets benefited equally from Consistency.

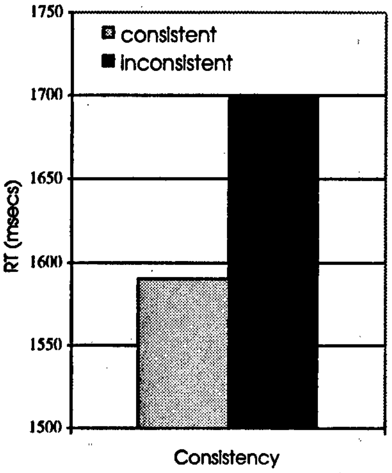


Figure 7. Space-to-space results.

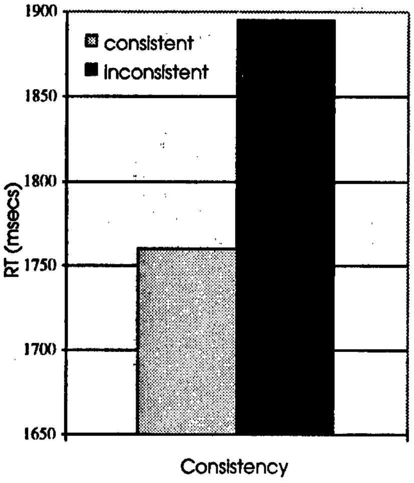


Figure 8. Time-to-time results.

Cross-domain schema consistency

Space-to-time: See Figure 9. Subjects responded significantly faster to Consistently presented targets (mean RT = 1973 msecs), than to Inconsistently presented targets (mean RT = 2088 msecs) when spatial prime questions preceded temporal target questions (F=4.20, p<.05). This schema-consistency effect means that there was transfer from the domain of space to the domain of time. This finding corroborates the hypothesis that people use spatial schemas to think about time. There was no effect of Target type, and no interaction between Target type and Consistency. This means that both ego-moving and time-moving targets benefited equally from Consistency.

Time-to-space: See Figure 10. There was no effect of Consistency in this condition (F=.71, p=.4). Response time to Consistently presented targets (mean RT = 1562 msecs) did not differ at all from response times to Inconsistently presented targets (mean RT = 1571 msecs). This means that there was no transfer from the domain of time to the domain of space. There was no interaction between Target type and Consistency, meaning that the ego- and object-moving targets were equally unaffected by Consistency.

These results are consistent with the hypothesis that temporal scenarios are understood and structured in terms of on-line mappings from the domain of space. These findings are also consistent with the results of Experiments 1 and 1a. These results confirm that the domains of space and time share structured relational information on-line. Furthermore, we found that the transfer is directional; there is an effect of schema consistency when the transfer is from space to time, but not the reverse.

DISCUSSION

Consistent with the metaphorical representation hypothesis, we find an asymmetry in the sharing of relational information between the conceptual domains of space and time. There was an effect of schema-consistency when temporal targets were preceded by spatial primes: subjects were slowed in solving problems about temporal relations if they had just completed schema-inconsistent problems about spatial relations. There was no effect of schema-consistency when spatial targets were preceded by temporal primes: subjects were not slowed in solving problems about spatial relations if they had just completed schema-inconsistent problems about temporal relations. These findings support the metaphori-

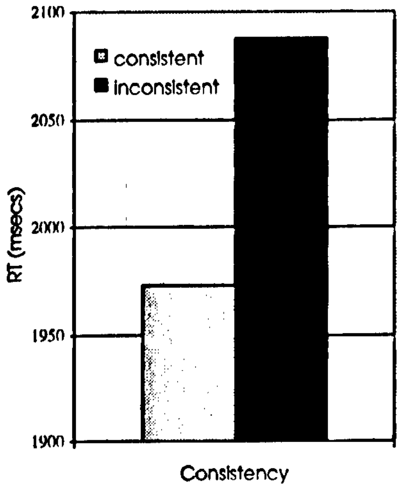


Figure 9. Space-to-time results.

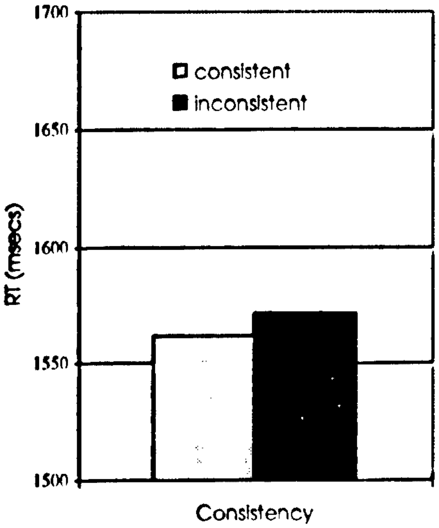


Figure 10. Time-to-space results.

cal representation hypothesis. There appears to be directional, on-line transfer of information from the concrete domain of space to the abstract domain of time. Results described above disconfirm the alternative hypothesis that the conceptual domains of space and time share generic domain-independent relational schemas. Ego-moving and object-moving schemas appear to be imported (borrowed) on-line from the domain of space, and used to organize events in time.

CONCLUSIONS

We found that people understand time in terms of space, but not space in terms of time. In Experiment 1, subjects were influenced by spatial perspective when reasoning about events in time. In Experiment 2, we showed that subjects were slowed in processing temporal statements if they were primed with an inconsistent spatial schema. This effect of consistency was present only in transfer from space to time, and not from time to space, indicating that there is a directional structure-mapping between these two domains. These findings lend support to the metaphorical theory of representation. It appears that abstract domains such as time are indeed structured on-line as metaphorical mappings from more concrete and experiential domains such as space.

It is still unclear, however, whether linguistic metaphors shape the way we think about abstract domains, or whether they simply reflect pre-existing conceptual mappings. A set of cross-linguistic studies is currently underway examining the role language in shaping abstract thought.

ACKNOWLEDGMENTS

The research presented in this paper was supported by a Graduate Research Fellowship from the National Science Foundation. This research program has greatly benefited from discussions with Dr. Gordon Bower, Dr. Herbert Clark, Dr. Dedre Gentner, and Dr. Barbara Tversky. Portions of this research have been presented at the 19th annual meeting of the Cognitive Science Society at Stanford University.

REFERENCES

Bennett, D. C. (1975). *Spatial and temporal uses of English prepositions: an essay in stratificational semantics*. London: Longman Group.

Boroditsky, L. (1997). Evidence for metaphoric representation: Perspective in Space and Time. In the *Proceeding of the 19th annual Meeting of the Cognitive Science Society.*

Clark, H.H. (1973). Space, time, semantics, and the child. In T.E. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press.

Fillmore, C. J. (1971). *The Santa Cruz lectures on deixis.* Bloomington, IN: Indiana University Linguistic Club.

Gentner, D., Imai, M., & Boroditsky, L. (in preparation). As time goes by: Understanding time as spatial metaphor.

Glucksberg, S., Brown, M., & McGlone, M.S. (1993). Conceptual analogies are not automatically accessed during idiom comprehension. *Memory and Cognition, 21,* 711-719.

Lakoff, G. & Johnson, M. (1980). Metaphors we live by. Chicago: University of Chicago Press.

Lakoff, G. & Kovecses, Z. (1987). The cognitive model of anger inherent in American English, In *Cultural models in language & thought*, (pp. 195-221).

Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics, 1*, 207-246.

Murphy, G. (1996). On metaphoric representation. *Cognition, 60,* 173-204.

Traugott, E. C. (1978). On the expression of spatio-temporal relations in language. In J. H. Greenberg (Ed.), *Universals of human language: Vol. 3. Word structure* (pp.369-400). Stanford, California: Stanford University Press.

METAPHOR: SHARED EXPERIENCE STRUCTURE OR CROSS-DOMAIN ANALOGY?

Maciej Haman
University of Warsaw
Faculty of Psychology
Stawki 5/7
00-183 Warszawa, Poland
e-mail: meh@sci.psych.uw.edu.pl

ABSTRACT

There is a substantial controversy concerning the role of metaphor in conceptual structure. According to one view (e.g. Lakoff and Johnson, 1980, Gibbs, 1996) some basic, direct experiences are used to structure and conceptualize, by the mean of metaphor, other, more abstract and not directly experienced matter. Another view (e.g. Keil, 1986, Gentner, 1989, Murphy, 1996a) treats conceptual metaphors as a kind of cross-domain structural analogies which could be constructed only on the basis of pre-existing conceptual knowledge in both domains. Some empirical evidence based on the study of priming in complex metaphor comprehension is presented for either position, however deeper theoretical analysis favours structural analogy hypothesis.

INTRODUCTION

Overview of metaphoric representation hypothesis

There is a lot of work done, mostly within cognitive linguistics, to demonstrate that abstract concepts like social engagements (love) and position, emotion, or mental states are represented metaphorically (later in this paper "MRH" abbreviation stands for metaphoric representation hypothesis). The idea is that image-schemata of bodily experiences, such as *orientation*, *containment* etc. provide both structure and, at least partly, meaning for more abstract concepts such as *love*, *anger*, *argument* or *social position* (Lakoff and Johnson, 1980; Johnson, 1987, Lakoff, 1987, 1991; Gibbs, 1992, 1994). The evidence for that claim originates mostly from linguistic data: the examination of frequency and commonalties between languages of the use of idiomatic expressions, and on soundness of this kind of metaphor. The common example is LIFE IS A JOURNEY metaphor, where *journey* relies closely on experiencing moving, and *life* is an abstract concept.

The hypothesis however was never taken seriously under examination in psychology (Baranski, 1996; Murphy, 1996a). Gibbs (1996) in his discussion with Murphy, cites several psychological data as supporting the metaphoric concepts hypothesis, but the evidence seems to be indirect at best. On the other hand Murphy's (1996a, b) criticise linguistic evidence, and claims that the universality of some idiomatic expressions could be better explained by appealing to the structural similarity hypothesis. That claim seems to be much better nested in the experimental data (Gentner, 1989; Medin, Goldstone, and Gentner, 1993).

Structural similarity hypothesis

Structural similarity hypothesis assumes that analogies, metaphors and structural similes based on systematicity of the mapping between base and target domains. The systematicity is estimated over the size and structure of the relational system matched between domains, with higher order relation (e.g. causal ones) weighted higher than first order relations (e.g. *bigger than*) and still higher than object attributes (e.g. perceptual properties like colour or size).

Developmental consequences of the MRH

On our part we would like to note that, when considered within psychological theory, MRH is, first of all, developmental hypothesis. The crucial test for it is then to demonstrate that (1) schematic representation of bodily experiences precedes acquisition of abstract concepts, and (2) children use the structure of their bodily experiences to acquire abstract concepts. No one however tested these hypotheses directly. The analysis of contemporary developmental studies gives no evidence either for the second claim, nor for the first, which intuitively seems to be much more plausible. Although development of abstract conceptions such as concepts of mental activities, or social relations, is perhaps well grounded in direct (mostly perceptual, but also motor) experiences, the paths of development seems to be separate at very early stages (Baron-Cohen, 1995; Leslie, 1994; Premack and Premack, 1995; but see also Smith and Katz, 1996, for alternative view). Also many early analogies in young children, even if superficially similar to Lakoff and Johnson's orientational metaphor are more likely to be based on structural similarity (Gentner and Ratterman, 1991; Gentner et al., 1995).

Reanalyses of the studies of early use of metaphor, which were not directly designed to test MRH, gives no clear support for it too. For example in our study of the pre-schoolers' understanding of physical transfer metaphor for mental actions (Haman, 1991, 1997) we found the pattern of answers which could be better explained by structural similarity hypothesis than MRH. Children between four and seven asked to interpret such expressions as "to give an idea" or "thoughts scattered" in the context of their play and school activities demonstrated a developmental shift from mixed to mental interpretation. While younger children correctly interpreted "to give an idea" as "to tell it out", they also incorrectly inferred that only the recipient will have the idea when given. Older children correctly interpreted both parts of the metaphor. As far this result could be an evidence for the hypothesis that abstract representation of mental transfer is build step by step on the physical transfer metaphor. However we have found very few purely literal interpretations even in youngest children. Moreover in nonmetaphoric condition children asked exactly the same question "who has the idea", easily realised that both agent and recipient have it (or sometime even attributed the "copyrights" to the agent only). Metaphor played here misleading rather than constructive role. Second, the "U"-shaped change was observed in the metaphor understanding task. There was an intermediate level of performance, at which children could realise that both agent and recipient have the idea, but failed to find "to give" = "to tell-out" equivalence. This pattern of results suggests, that there was a change in the structure of children understanding of mental transfer, which promoted new level of structure mapping between tenor and vehicle. That change seems to be better linked with parallel development of theory of mind, rather than to be lead by physical transfer metaphor. We assume then that the study of cross-domain structure mapping and transfer is the important part of empirical exploration of MRH vs. structural similarity trade-off.

DOMAIN SPECIFICITY AND METAPHOR

There are several studies demonstrating that metaphor understanding proceeds domain by domain rather than single term to term (Keil, 1986; Kelly and Keil, 1987; Tourengeau and Sternberg, 1982). To some extent it is also congruent with MRH. However, contrary to MRH it was found that some fair level of understanding and structurization in both domains (tenor and vehicle) was required to establish a metaphoric relation (Keil, 1986). Note, that at least some of the metaphors used in Keil's study matched the type of metaphors considered in MRH. We are going now to discuss domain specific conceptual representations and their links to MRH vs. structural similarity trade-off.

Domain specificity

There is a good deal of work elaborating the hypothesis that concepts are represented within larger structures: domains (Keil, 1986, 1989; Hirschfeldt and Gelman, 1994a; Haman, 1997a). It is hard to provide a single and exhaustive definition of the domain (see Hirschfeldt and Gelman, 1994b, and other papers in the Hirschfeldt and Gelman, 1994a, volume; Haman, 1997a, and b). For the current problem two properties of domain are the most important ones: quasi-modular status, and underlying domain theories. First of them conflicts with developmental interpretation of MRH. Second one provides some interesting consequences for metaphorical asymmetry, which was claimed to support MRH.

Foundational domains

There is no agreement how many domains are there, and which of them are developmentally foundational. Most of the researchers agree however that the physical object/mental entity distinction is based on very early cognitive achievements. There is no place to discuss this issue here. While we are aware that there is a lot of arbitrariness in that, we think that it is justified to assume that at least adults (but perhaps already children at preschool age) represent mental and social phenomena, artefacts, inanimate natural kinds, plants, and animals (as well as some nominal kinds like language and number) in separate domains. These domains differ on the dimension of complexity and consistency of their foundational theories, from complex and consistent domain of animals through plants and inanimates, to artefacts, which lack specific causal theories[1].

Metaphorical asymmetry

One of the problems which Gibbs (1996) raises against structural similarity hypothesis is the asymmetry of foundational metaphors. People tend, for example, to speak about love in terms of the trip, but not vice-versa. That asymmetry isn't however exceptionless. In Polish for example you could say: "On kipi z wsciekloscі" (*He just boils of furry*), which is classical example of pressure in the container metaphor for emotions, however the reverse "Zupa w garnku kipi wsciekle" (*Soup in the pot boils furiously*) is almost equally sound. Indeed, Murphy's (1996a, b) appeal to typicality and salience to explain that asymmetry is not convincing.

Our studies reported in Haman (1997) allowed however to propose another explanation. We have adopted Keil's (1989) hypothesis, that conceptual domains differ on the dimension of representational complexity at the domain's theory level (in general natural kinds, and especially animate, are underlaid by complex causal network, and that complexity declines trough inanimates, complex artefacts, simple artefacts to nominal kinds, while the arbitrariness and well-defindness raises in that direction).

Using metaphor understanding and *ad hoc* categorization tasks we have show that domains which are not underlaid with complex causal theory are good "exporters" of objects and structures to other domains (e.g. good metaphor vehicles), while domains with riche theories tends rather to assimilate elements of other domains (are good "importers", or metaphor tenors).

It is not necessarily obvious if thinking about structure mapping as a process realised by special device like SME in Gentner's (1989) model. SME first generates partial mappings and then makes a decision on the bases of maximum systematicity. On contrary we think about structure mapping as a process of establishing cross-domain links and finding correspondences in conceptual structure *in situ*. Highly interconnected and coherent domains are much more likely not to generate a link to less structured domain, could however easier find correspondences to the structure "imported" from other, less structured, domain. The systematicity principle plays here the same role as in Gentner's model, but it is computed here in the context of the entire structure of the target domain in the problem.

Rationales for the current study

No one of the studies noted above was designed directly to test MRH. Here we are going to propose a method that explicitly contrasts metaphoric representation hypothesis with cross-domain structure mapping. We have designed two experiments as a preliminary test, which could be a starting point for more extensive research. This is still a study of adults' metaphor comprehension, so it leaves developmental issues still unexplored.

EXPERIMENTS

Overview

The method of our experiments is based on Boronat and Gentner (1990) compound (extended) metaphor understanding study. Their materials consisted of metaphors composed of two parts, each of which was a legal and complete metaphor itself. Both parts could be either consistent, i.e. their vehicles originated from the same conceptual domain, either inconsistent - vehicles originated from different domains. In both cases, however, the interpretation of both parts taken together provided a coherent meaning for entire utterance. Consistent metaphor example is: *Was Anna still boiling mad when you saw her? — No, she was doing a slow simmer.* Inconsistent metaphor could be: *Was Anna still raging beast when you saw her? — No, she was doing a slow simmer.* Both metaphors have mutually same interpretation, the very same final component, and very similar structure, however the initial component of the inconsistent metaphor use animals domain as a vehicle, while in initial component of consistent metaphor, as well as of final component take fluids dynamics as vehicle.

The main idea of the experiment was that if the metaphors are processed domain by domain, first part of the metaphor will prime the understanding of the second part only if the metaphor was consistent^2.

In our study we have assumed that if MRH is correct, then compound metaphors consistent in respect to basic bodily experience will show stronger priming effect than metaphors based on domain consistency, as they access foundational conceptual relations. If cross-domain structure mapping hypothesis is correct, then domain knowledge and structure will support priming. Two experiments, described below differs, apart of the same general design, in the kinds of MRH metaphors explored, conceptual domain definition, and the way the priming effect was assessed. In the experiment 1. more general domain conception was adopted (based on Keil's, 1989, natural, artefact, and nominal kinds distinction), and only ontological metaphors (in Lakoff and Johnson, 1980, sense) were used. In Experiment 2. we introduced more fine distinctions on both dimensions.

EXPERIMENT 1.

Method

Subjects. The initial sample of 24 undergraduates paid for participation were tested. Nine of them were excluded from the final analysis because of extreme variance in their results (in respect to 2SD threshold criterion), so the final simple consisted of 15 subjects.

Materials. A set of component metaphors was created. Each of the components (simple metaphor) could be classified into one of the Lakoff and Johnson (1980) schemes (part/whole, container, and path/journey) and Keil's (1989) kind type (natural, artefact, and nominal). In order to combine any domain with any image-scheme, the total of 36 compound metaphors were created, so each compound metaphor was either consistent both on Lakoff and Johnson's (MRH consistency), and on Keil's dimension (DSC - domain structure consistency), either on one of them, or inconsistent at all. Compound metaphors were distributed over three equivalent sets of 12 elements. Five subjects from the final sample proceeded with each set. Each set consisted of three compound metaphors of each type: MRH+/DSC+, MRH+/DSC-, MRH-/DSC+, and MRH-/DSC-, where "+" means metaphor consistent, and "-" inconsistent at a given dimension.

Initial and final components of each metaphor were matched in respect to sentence length and complexity.

The example is: *Company is a complex puzzle composed of many small elements. But only such a complex team could reach the world-cup* (MTH+ based on whole/part metaphor, DSC-).

Procedure. Subjects were tested individually. Each subject read an instruction on the computer screen. The instruction asked subject first to type a space to display the initial component of the metaphor, read it, type the space again, and read the final component, then type the space when ready to explain the entire (compound) metaphorical sentence. Only single component sentence was displayed at time. First touch of the space key settled the clock on and the second off, so for each compound metaphor we have got 2 reaction times: one for the first component and one for the second. Test trials were preceded with single training trial. Subjects' explanations were presented verbally and tape-recorded, (however we will not refer to them in the results' discussion).

Scoring. The priming effect was assessed by estimating per-cent of the time necessary to explain the entire metaphor after reading second component in comparison to the first component. That was expressed in the equation: (RT1*100)/RT2, where RT1 is a time of reading initial component and RT2 is time necessary to explain the metaphor after the final component was displayed. Finally for each subject the mean of three metaphors representing the same configuration of consistency was computed.

Results and discussion.

Table 1. summarize the results over all three metaphor sets. 3x2x2 ANOVA (set by MRH consistency by DSC, with repeated measures within both consistency factors) was computed. There were main effects of MRH (F[1;12]=5.61, p<.035), and of metaphor set (F[2;12]=7.65, p<.007), and DSC effect at tendency level (F[1;12]=3.19, p=.10). No interaction effect was found.

	DSC+	DSC-	Mean
MTH+	155.13	206.98	180.56
MTH-	191.23	266.08	228.66
Mean	172.68	236.94	204.61

Table 1. Mean times necessary to understand metaphor as a per-cent of RTs for the first component

Significant main effect of MTH and only tendency in the direction of DSC seems to support metaphoric representation hypothesis as a main factor in metaphor processing. Close look on data make that interpretation implausible. Taken together, as well separately in each set, MTH-/DSC+ metaphors were processed faster than MTH+/DSC-. The difference is not very large, and so not significant, but systematic. So the reason for not significant DSC effect is a higher variance within that category. We could try to explain the source of that variance.

The most important sources of error variance (and perhaps that which caused exclusion of 9 subjects) were troubles to establish full equivalence and relative soundness of metaphor components. Metaphors' soundness varied across metaphors and sets, as documented by significant set effect (however the general pattern of results was similar across sets - there was no interaction between set and other factors). For example it is not easy to find good Lakoffian type metaphor based on inanimate natural kinds. As we have used very broad domain concept some domain consistent metaphors have in fact very little common ground. To avoid part of these problems we have planned experiment 2.

EXPERIMENT 2.

The aim of experiment 2. was to minimize variance related to procedure. We have constructed new material as well as new measure of priming effect. Finer distinctions of domain and different types of Lakoffian metaphors allowed us to asses additionally the role of specific domain and metaphor type. We have also altered the method of manipulation and measuring priming. Here we vary only initial component (primer). Final component is the same for MRH+, DSC+, and inconsistent metaphors.

Method

Subjects. Twenty-four undergraduates paid for participation. Three additional subjects were excluded because of lacking results.

Materials. The basic set of 12 final components was created. For each of them there were 3 different initial components: MRH consistent (MRH+), domain consistent (DSC+), and inconsistent in either domain or MRH (INC), so the product was 36 compound metaphors divided into 3 experimental sets. The metaphors with the same final component were never included into the same set. As in the first study, the compound metaphor as a whole had a congruent meaning independently of MRH or DSC consistency.

Lakoff and Johnson differentiate several types of metaphoric conceptualisation. According to that MRH+ metaphors were farther classified as structural, containment, and orientational. DSC+ metaphors' vehicle belonged to one of three domains: inanimate natural kinds, plants, and animals. To achieve maximum consistency of results all MRH+ final components were based on artefact domain (we will discuss that later, in results' section). Each experimental set consisted of four MRH+ metaphors (one of each type with one type doubled), four DSC+ metaphors (two inanimate natural kind, and two animate: one plant and one animal), and four inconsistent (INC) metaphors. An example is: *His hopes had been like towers* (MRH+ based on orientational metaphor: up=better, down=worse) or *His hopes were like galloping bizons* (DSC+ based on the domain of animals). or *His hopes were like attacking tank division* (INC), with the common final component: *After some time his hopes became to be like a little bird dropped from the nest.*

Procedure. Procedure was fully analogical to that in experiment 1.

Scoring. As the same final components were used for different consistency models, we have used only RT2, i.e. time necessary to explain the metaphor after the second component was displayed, as a measure of priming effect. If for a single subject were more than one RTs in the same cell of ANOVA design, an average was computed (e.g. for two DSC+ metaphors based on inanimate).

Results and discussion.

MRH+	7.184
DSC+	8.269
INC	9.133
Mean	**8.195**

Table 2. mean RTs for three types of metaphor consistency

Table 2. contains summary results for consistency type. 3x3 ANOVA (experimental set by consistency, with repeated measure within consistency) was performed. There was no effect of experimental set, nor set by consistency interaction. The main effect of consistency was highly significant (F[2;45]=7.45, p<.0015). Planned comparisons revealed that both MRH+ and DSC+ metaphors were faster processed than inconsistent (INC): F[1;21]=12.60, p<.002 and F[1;21]=5.14, p<.035 respectively. MRH+ metaphors were also processed faster than DSC+, although the difference only approach tendency level (F[1;21]=3.92, p=.061). Thus again we have got an ultimate evidence for both sources of metaphor consistency. Relatively faster performance in the case of MRH+ metaphors than DSC+ is reasonable, as all MRH+ metaphors had artefact domain as vehicle in the initial component. We have argued earlier, that artefacts are very good vehicles, as they are not underlaid by systematic causal-explanatory network.

We have performed also two ANOVAs to prove the Lakoffian metaphor type (structural, orientational, container) and domain (inanimates, plants, animals) effects. 3x2x3 design (experimental set by MRH+/DSC+ by metaphor type) gave no significant effect or interaction. Second 3x2x3 design (experimental set by MRH+/DSC+ by domain) also gave no significant effect, there were however some interesting tendencies. Plant metaphors were processed faster than animal metaphors (F[1;21]=3.46, p=.077). The overall effect of domain also approach tendency level of significance (F[4;42]=2.25, p=.115). The results are congruent with our expectation, that metaphors based on animal domain need more time for processing because of complex net of underlying relations.

GENERAL DISCUSSION AND CONCLUSION

We have searched for empirical test to deal with metaphoric representation vs. structural similarity hypothesis trade-off. Priming in compound metaphor understanding task could provide some date to estimate the role of relations proposed by MRH and domain-specific naive theories in providing a common ground for both components of the metaphor. As far the results provide some support for both hypotheses. There are however some not yet proved arguments for domain knowledge view. First, domain of the metaphor vehicle seems to influence also processing of MRH+ metaphors. Second, we have not found any evidence for privileged position of MRH+ consistency, as could be expected if it is a base for conceptual representations. As Murphy (1996a, b) argues, the MRH+ consistency also could be explained by structure similarity, and the recurrent question is what is developmentally earlier. It is however very hard to test. It is important then to search for the evidence also in adult performance.

Our study was designed as a pilot attempt to approach the problem experimentally. We think that the results give reasons to master the compound metaphor task in order to independently control consistency, domains, and metaphor soundness.

ACKNOWLEDGEMENTS

The research was supported to Maciej Haman by grant no. BST 541/56 from University of Warsaw, and grant no. 1 P106 021 06 from Commission for Scientific Research.

Boicho Kokinov, Dedre Gentner, and Keith Holyoak gave me the reason to write this paper, inviting me to participate in the workshop. Michal Baranski participated in the initial problem formulation, and had a good deal in designing and running the experiment 1. as a part of the requirement for his Master degree. Special thanks for Robert Sliwinski who invented and ran experiment 2., and his defence of Lakoff and Johnson allowed me state better my position.

REFERENCES

Baranski, M. (1996). *Pojeciowe zródla spojnosci metafor* [Conceptual sources of metaphor coherence]. Unpublished master thesis. University of Warsaw.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA.: The MIT Press.

Boronat, C. B., & Gentner, D. (1990). *Effects of base shift and frequency in extended metaphor processing.* Unpublished manuscript, University of Illinois, Urbana-Champaign.

Gentner, D. (1989). The mechanism of analogical learning. In: S. Vosniadou & A. Ortony [Eds.]. *Similarity and analogical reasoning.* Cambridge: Cambridge University Press.

Gentner, D. & Ratterman, M.-J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes [Eds.], *Perspectives on language and thought: In-*

terrelations in development. Cambridge: Cambridge University Press.

Gentner, D., Ratterman, M.-J., Markman, A., & Kotovsky, L. (1995). Two forces in the development of relational similarity. In T. J. Simon & G. S. Halford [Eds.], Developing cognitive competence: New approaches to process modelling. Hillsdale, NJ: Erlbaum.

Gopnik A., & Wellman, H. M. (1994). The theory theory. In: L. A. Hirschfeldt & S. A. Gelman [Eds.], *Mapping the mind.* Cambridge: Cambridge University Press.

Gibbs, R. W., Jr. (1996). Why many concepts are metaphorical. *Cognition, 61,* 309-319.

Haman, M. (1991). *Wyodrebnianie pojec przyczynowo-wyjasniajacych z schematow dzialania* [The emergence of causal-explanatory concepts from action schemata]. Unpublished PhD Dissertation. University of Warsaw.

Haman, M. (1997a). Domain specificity and cross-domain transfer. *Psychology of Language and Communication, 1,* no. 2, 69-84.

Haman, M. (1997b). Concepts in the potboiler: The final debate. *Psychology of Language and Communication, 1,* no. 2, 69-84.

Hirschfeldt, L., & Gelman, S. A. [Eds.] (1994a). *Mapping the mind.* Cambridge: Cambridge University Press.

Hirschfeldt, L. A., & Gelman, S. A. (1994b). Toward a topography of mind: An introduction to domain specificity. In: L. A. Hirschfeldt & S. A. Gelman [Eds.], *Mapping the mind.* Cambridge: Cambridge University Press.

Johnson, M. (1987). *The Body in the Mind.* Chicago: University of Chicago Press.

Keil, F. C. (1986).Conceptual domains and the acquisition of metaphor. *Cognitive Development, 1,* 73-96.

Keil, F. C. (1989) *Concepts, Kinds, and Cognitive Development.* Cambridge, MA: The MIT Press.

Kelly, M. & Keil, F. C. (1987). Conceptual domains and the comprehension of metaphor. *Metaphor and symbolic activity, 2* , 33-51.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lakoff, G. (1987). *Women, fire, and dangerous things.* Chicago: University of Chicago Press.

Lakoff, G. (1990). The invariance hypothesis: Is abstract reason based on image-schemes? *Cognitive Linguistics, 1,* 39-74.

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In: L. A. Hirschfeldt & S. A. Gelman [Eds.], *Mapping the mind.* Cambridge: Cambridge University Press.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254-278.

Murphy, G. L. (1996a). On metaphoric representation. *Cognition, 60,* 173-204.

Murphy, G. L. (1996a). Reasons to doubt the present evidence for metaphoric representation. *Cognition, 62,* 99-108.

Premack, D., & Premack, A. J. (1995). Origins of human social competence. In: M. S. Gazzaniga [Ed.], *The cognitive neurosciences* (pp. 205-218). Cambridge, MA: The MIT Press.

Smith, L. B., & Katz, D. B. (1996). Activity-dependent processes in Perceptual and Cognitive Development. In R. Gelman & T. K.-F. Au, *Perceptual and cognitive development.* New York: Academic Press.

Tourangeau R., & Sternberg, R. J. (1982). Understanding and appreciating metaphors. *Cognition, 11,* 203-244.

---

[1] The position of mental/social domain on this continuum is not clear. While it is well documented that even preschoolers have well elaborated theories of mind (see e.g. Gopnik and Wellman, 1994), then many social categories and relations seems to be highly sterotypic nad idiosyncratic, allowing concurrent representations of foundamentally inconsitent believes.

[2] No priming is expected if the metaphors are processed at tle level of single concepts, and the meaning of the compound metaphore is a sum of component interpretations.

ANALOGY UNDERLIES SENTENCE GENERATION AND INTERPRETATION

Nili Mandelblit
Langage et Cognition
LIMIS, CNRS
BP 133, F91403
Orsay Cedex, FRANCE
mandelbl@cogsci.ucsd.edu

ABSTRACT

We argue that the generation of every sentence involves first the perception of analogy between two conceptual structures, and then an operation of linguistic mapping. Sentence interpretation starts with an attempt to reconstruct the analogical mapping configuration underlying the sentence (i.e., the mapping operation performed by the speaker).

According to this view, an important role of *grammar* is to formally mark various analogical mapping configurations, thereby providing cues to the hearer in the interpretation process. Different grammatical systems have evolved in different languages to formally mark such mapping operations.

1. WORKING ASSUMPTION: THE CONSTRUCTION GRAMMAR VIEW

The basic assumption in the analysis is the Construction Grammar view (as proposed by Fillmore & Kay, 1993, as well as Lakoff, 1987, Goldberg, 1995, and others), the basic propositions of which are also shared by Langacker's Cognitive Grammar approach (1987, 1991). The assumption is that languages are made up of *constructions* - pairings of grammatical forms (syntactic or morphological) and semantic structures. Mastery of language consists of mastery of these form-meaning pairs. Syntactic forms in particular are associated with conceptual schemas representing *generic event structures* which are basic to human experience, such as manipulation of objects, bodily movement through space, and dynamics of force and enablement (Goldberg, 1995). These schemas are thought of as tools for organizing comprehension and communication and can structure (indefinitely) many perceptions, images, and events (see also the notions of *image schemas* and *conceptual archetypes* in Johnson, 1987, Langacker, 1991, Talmy, 1988, Turner, 1996). In recent years, cognitive scientists have found strong evidence for the existence of such event schemas. Examples include the role of event schemas in metaphorical understanding (Lakoff & Johnson, 1980, Sweetser, 1990), and as precursors for language acquisition by children (Mandler, 1992, In press).

Given the Construction Grammar assumption (and its cognitive linguistics extensions), we can now talk about linguistic entities such as the English Transitive Construction. The syntactic form of the construction is [NP V NP] (==SUB V OBJ). Its associated semantic schema is the archetypal "transitive" event (as defined, for example, in Givón, 1984): an agent (typically human), who volitionally acts on (i.e., exerts physical force on) and affects another entity (a patient)[1]. Each role in the semantic

---

[1] This schematic event structure clearly represents only the most prototypical sense of the simple Transitive construction. A full description of this grammatical construction involves a network of extensions to the prototypical sense as well as a list of idiomatic uses of the construction (as, for example, in Goldberg's study of constructions, 1995). The network description of a construction is analogous to a description of a prototypical sense of a lexical item which nearly always involves a network of polysemous and metaphorical extensions.

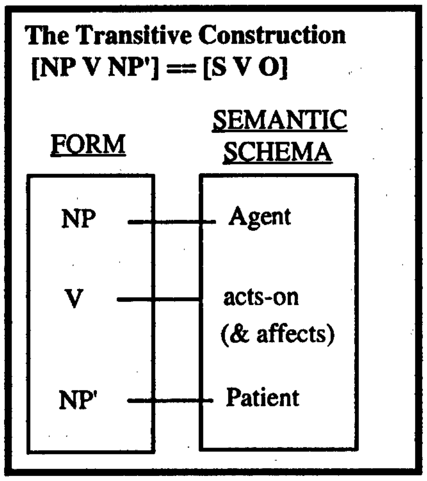


Figure 1. The English Transitive Construction.

schema is associated with one grammatical category in the syntactic pattern (Figure 1): the agent role is associated with the subject NP, the patient role with the object NP, and the force-dynamic relation between the two entities is associated with the main verbal slot. The semantic schema and its association with the syntactic form are extracted by speakers from frequently encountered instances of the construction (i.e., instances of two-participant transitive sentences).

2. ANALOGICAL PERCEPTION AND MAPPING OPERATIONS IN SENTENCE GENERATION.

Consider a basic transitive sentence in English, such as "Mary poisoned her lover" (generated, say, by a detective investigating a murder case). The actual event in the world involved a complex sequence of events: Mary first made a (probably intentional) decision to kill her lover. She decided to use poison. She found (or bought) some poison and put it in her lover's food. The lover ate it, felt sick, and after a while died.

But at some more abstract level, this sequence of events is also perceived by the speaker as an instance of a more generic event structure: An agent (Mary) acting on and affecting a patient (her lover). At this abstract level, the actual details of the event are ignored, and an *analogy* is *perceived* between the high-level structure of the novel event and the "transitive" event schema ('Agent act-on and affects Patient'). At this level of abstraction, Mary who initiated the whole causal sequence of events is perceived as analogous to (or an instance of) the Agent role in the generic transitive event schema (while ignoring other intermediate causal forces involved in the event). The lover — the salient affected entity in the causal event sequence — is perceived as an instance of the Patient role in the transitive event schema (while ignoring other less salient affected objects, such as the poison and the food manipulated by Mary as well).

The perception of the analogy between the high-level structure of the conceived novel event and the structure of the transitive event schema motivates the speaker to use the transitive *syntactic construction* [NP V NP] (associated with the transitive schema, Fig. 1.) as a linguistic *integrating* frame (Fauconnier & Turner, in press) for communicating the event that occurred in the world. The perceived analogy between Mary and the Agent role in the transitive event schema leads to the *linguistic* association (or *binding*) of the lexical item 'Mary' (that represents the person Mary) with the subject NP slot in the Transitive syntactic construction (that represents the Agent role in the transitive event schema). Likewise, the perceived analogy between the murder victim (the lover) and the Patient role in the transitive event schema leads to the linguistic association of the phrase 'her lover' (representing the affected entity) with the object NP slot in the Transitive syntactic construction (representing the Patient role in the transitive event schema).

The speaker now also has to choose which aspect of the conceived event to express through the verbal slot of the Transitive construction. In the sentence 'Mary poisoned her lover', the lexical item 'poison' denotes the *substance* Mary used to affect (kill) her lover. Note that this lexical item ('poison') represents only one aspect in the complex event, but this aspect is considered central (or salient) enough to be used as a linguistic representative of the whole event, and as a trigger in the hearer's mind for reconstruction of the whole causal event sequence. Figure 2 illustrates the analogical mapping operation between the two conceptual structures: the structure of the rich conceived event (which is composed of a sequence of temporally and causally related sub-events) and the structure of the Transitive event schema. The *linguistic* binding of lexical items (and their phonological form) with syntactic slots in the transitive construction follows the perceived conceptual analogy and the mapping operation across the two conceptual structures. This linguistic binding constitutes the basic operation for sentence formation (leading to the string: 'Mary poisoned her lover').

Note that we did not represent in Figure 2 the "generic space" (Fauconnier & Turner, 1994), which reflects the common structure and organization shared by the two input structures (the conceived causal event and the Transitive construction). It is by virtue of this common abstract structure that analogy can be perceived and mapping performed across the two input structures. The generic structure in Figure 2 is the transitive event schema, representing both the semantics of the transitive syntactic construction, and that of the high-level abstracted structure of the conceived event.

To sum, what are the basic cognitive skills required for the generation of the sentence *Mary poisoned her lover* as a description of the actual complex conceived event? The discussion above suggests that the following minimal skills are required:

(1) The ability to **abstract** the representation of the rich conceived event in the world to a level where it shares structure and organization with a generic event schema (e.g., the transitive event schema - 'Agent act-on/affect a Patient'). This abstraction operation is not explicitly illustrated in Figure 2.

(2) The ability to perform the **structural mapping** between the two representations (an example of such mapping configuration is given in Figure 2).

(3) **Mastering of the conventional form-meaning associations in the language** both between syntactic constructions and event schemas (e.g. the association between the transitive syntactic pattern [NP V NP] and the transitive event schema, Figure 1), and between lexical-phonological items and entities or relations conceived in the world.

(4) The ability to perform the **linguistic binding** operation between lexical items and syntactic slots (in a syntactic construction) following both the perceived conceptual analogy (at the semantic level) and the grammatical conventions of the language (languages differ in the type of linguistic binding they permit, or prefer, and how they mark them, as will be discussed in the next section). This last operation is the basic operation underlying sentence (and probably discourse) generation.

The first two skills defined above are general **analogy making** skills (as discussed, with some variations in, for example, Hofstadter et al, 1995, Holyoak & Thagard, 1994, Indurkhya 1992; the first skill of abstraction parallels for example Hofstadter's notion of "essence-extraction", proposed to be the first stage in analogy making). The third and fourth skills are **lin-**

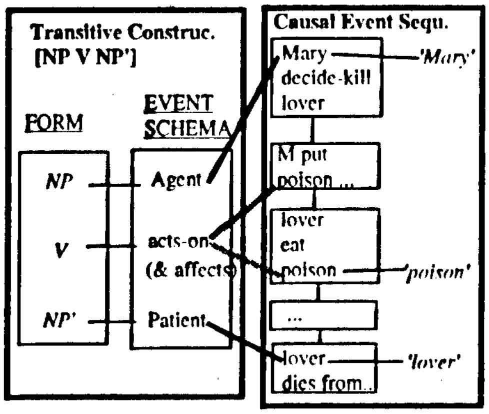


Figure 2.

guistic skills (the third skill, and in particular the details of the form-meaning associations in each language, has been a main topic of study in the Cognitive Linguistics literature). All four skills require the ability to **access** conceptual structures (linguistic and non-linguistic) in memory, and **map** (or **bind**) these structures onto one another [2].

3. MAPPING CONFIGURATIONS AND GRAMMAR

In the previous section we discussed one example of analogical mapping in sentence production. An analogy is first perceived between a conceptual abstraction of a complex conceived event in the world and the semantic structure of one of the language's syntactic constructions. This analogy leads to the linguistic expression of the event by means of the syntactic construction. The sentence generation operation is based on linguistic association of lexical items and syntactic slots in the construction, following the perceived conceptual analogy and mapping.

The conceptualization and communication of a complex conceived event as an instance of a simple event structure (e.g., the transitive event schema) has clear cognitive advantages. This process of *conceptual integration* (Fauconnier & Turner, in press) facilitates the conceptual manipulation and categorization of the event, and its storage in memory. It also enables easier communication (a simple, short sentence can trigger the whole event sequence in the hearer's mind). From a linguistics point of view, this process allows reusing a small set of grammatical forms (syntactic constructions) for the expression of infinite number of novel, complex events.

If this process is indeed so useful cognitively, then it would be only natural if formal grammatical systems would evolve to formally mark such analogical mapping operations in order to systematize and facilitate their communication. Research on grammatical mapping and integration suggests that this is indeed the case.

Consider, for example, the active-passive grammatical dichotomy found cross-linguistically. What this dichotomy really tells the hearer is which participant in the conceived event has been linguistically bound with (and expressed by) the subject slot of the integrating syntactic construction. The active form typically tells the hearer that an agent (a source of energy) in a conceived event has been bound onto the subject slot of the syntactic construction, and the passive form tells the hearer that a patient (an affected entity) has been mapped onto the subject slot of the syntactic construction.

Figure 3 illustrates the difference in *mapping configuration* underlying the active sentence *the dog is eating*, and the passive counterpart *the dog is eaten*. The active-passive verbal grammatical forms *(be V-ing* vs. *be V-en)* define the different mapping configurations, thereby providing the hearer with *instructions* on how to link (map) the partial information provided by the lexical items in the sentence ('dog', 'eat') to the actual structure of the communicated event.

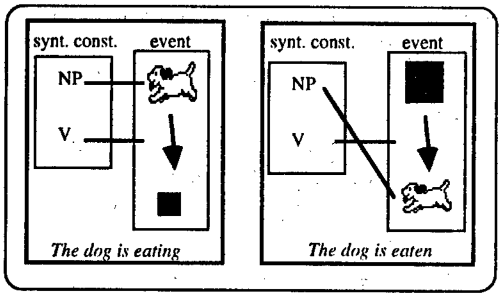


Figure 3. The active-passive mapping configurations (schematic description).

---

[2] For discussion of mapping and binding operations, see, for example, Fauconnier's study on *mapping in language and thought* (1997), Damasio (1989) and Sahstri et al, (1993) on binding and *convergence zones*, and Grush & Mandelblit (1998), Mandelblit & Zachar (1998), and Petitot (1995) on their interdisciplinary links.

In Mandelblit (1997, ms.), Hebrew verbal morphological constructions (*binyanim*) are analyzed in detailed. It is suggested that each morphological construction marks a particular type of *mapping configuration* between a conceived event (typically a causal sequence of events) and a syntactic construction. The morphological construction marks: (1) which participant in the conceived causal event (e.g., the causal force or the affected entity) has been mapped onto the subject slot of the syntactic construction (as in the active-passive contrast described in Figure 3); (2) which predicate in the conceived event (the causing or effected predicate) has been mapped onto the verbal slot of the syntactic construction. A summary of the mapping configurations associated with each of the seven principal morphological *binyanim* in Hebrew is given in Figure 4.

English, in contrast to Hebrew, does not possess a grammatical system as rich as the Hebrew morphological *binyanim* system to mark the link between the main verb in a sentence and the structure of the communicated event. For example, the verb 'ran' in *Mary ran around the block* and *Mary ran the dog around the block* looks exactly the same, even though in the first sentence 'ran' refers to the activity of the subject 'Mary' (the sole energy source of the running action), while in the second sentence 'ran' primarily refers to the activity of the patient 'the dog' (while Marry, who made the dog run, was not necessarily running herself). The verbal form 'ran' in both sentences denotes only a type of motion activity (of Mary or the dog), but not the relative role this activity plays within the general structure of the communicated event.

What types of linguistic mapping configurations from a conceived event onto a syntactic construction are found in English (where the semantics of the syntactic construction is taken to be analogous to an abstracted structure of the communicated event)?

Fauconnier & Turner (1996) analyze the mapping configurations underlying the use of the English Caused-Motion syntactic construction, studied by Goldberg (1995). The form of the English Caused-Motion construction is [NP V NP PP] (= SUB V OBJ OBL), and its associated semantic schema, as Goldberg suggests, is of a "caused motion" event ('X causes Y to move Z'). Examples of this construction include:

(1) The audience laughed the poor actor off the stage.

(2) Monica trotted the horse into the stable.

(3) The commander let the tank into the compound.

(4) Paul hammered the nail into the door.

In each of the sentences (1-4), a whole causal sequence of events [[X act] cause [Y move in direction Z]] is mapped (and conceptually integrated) into the caused-motion syntactic construction [NP V NP PP], based on perceived analogy between the abstract structure of the

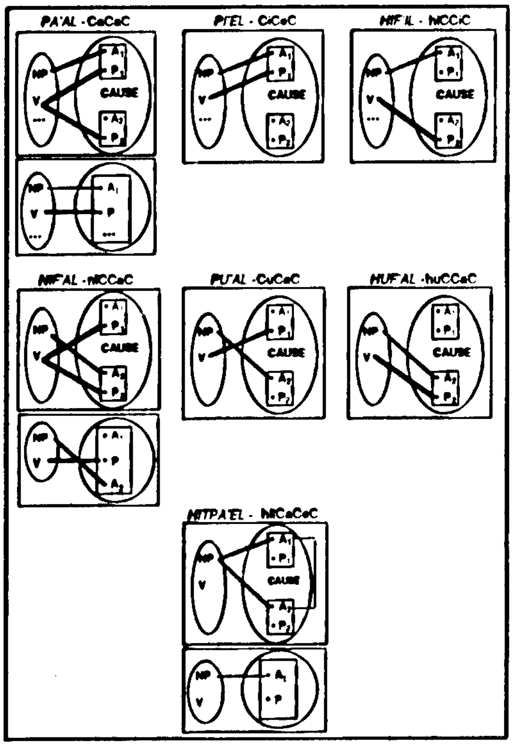


Figure 4. The mapping configurations marked by the different Hebrew verbal morphological binyanim (Mandelblit, 1997).

conceived event and the caused-motion semantic schema associated with the syntactic form.

In each sentence, different aspects of the conceived causal event sequence are mapped onto the verbal slot of the construction. In example (1), the verb *laugh* specifies the agent's causing action. In 2, the verb *trot* specifies the motion of the affected patient (the horse). In (3) the verb *let* does not specify neither the agent's causing action, nor the patient's motion, but rather the causal link (force dynamics) between the (unknown) commander's action and the tank's motion. In (4) the verb *hammer* specifies the tool used for achieving the caused-motion event. The last mapping (4) is most similar to the one observed in the first example discussed in this paper (*Mary poisoned her lover*), where the verb *poison* describes the means (substance) that the agent used to affect (kill) the patient [3]. Note that nothing in the English grammar marks to the hearer the mapping configuration underlying each sentence. It is up to the hearer to reconstruct the analogical links between the lexical information provided in the sentence and a probable conceived event in the world [4].

While it is possible to find in each language a basic similar set of conventional mapping configurations (either marked grammatically or not), languages seem to differ in which mapping configurations are 'favored' (used more often than others) in everyday speech (for implications of these differences to translation, see Mandelblit 1995, 1997).

But whatever the conventions are, speakers are able to come with novel surprising mappings, as exemplified in the following caused-motion sentence (from Fauconnier & Turner, 1996):

(5) The spy Houdinied the drums out of the compound.

The analogy in example 5 between the high-level structure of the conceived event and the semantics of the caused-motion schema (an analogy which led the speaker to express the conceived event through the caused-motion construction) is itself quite straightforward (as in examples 1-4). What makes example 5 look so creative is the unconventional underlying *mapping configuration*: the binding of 'Houdini' to the verbal slot, and what role Houdini plays in the conceived event . We will not go now into the details of the mapping (we leave it for the reader), but what examples such as 5 show is that the choice of a syntactic construction for expressing an event as a result of perceiving structural analogy between the event and the construction's semantics is just the first creative stage in sentence generation. Then, many different linguistic mappings may be used between the two analogous structures - some are entrenched, and often marked grammatically (as in the use of Hebrew *binyanim*, Figure 4), and others are completely novel and unpredictable (thereby requiring special effort during the process of interpretation). Current computational models of analogy and language processing can model the very entrenched linguistic mappings, but do not account yet for the real creative ones.

---

[3] The use of verbs such as *hammer* and *poison* in English has become so entrenched that today these verbs are viewed as denoting a whole causal event themselves rather than just the tool or substance used to achieve an effect. Note however that when these verbs first emerged in the language (through a so-called "verbal derivation" operation) they reflected a particular type of mapping configuration from events onto syntactic forms that speakers preferred to use. Similar new mapping configurations are still created everyday by speakers, and it is the goal of cognitive linguists to capture and describe this productive operation.

[4] Goldberg (1995:65) defines a hierarchy of possible relations between the semantics designated by a verb (V) and the semantics designated by the syntactic construction (C) it instantiates. By doing so, Goldberg defines in fact the various mapping configurations available in English between what the verb designates in the conceived event and the analogous semantics of the construction. The hierarchy Goldberg defines is as follow: 1. V is a subtype of C. 2. V designates the means of C. 3. V designates the result of C. 4. V designates a precondition of C. 5. (to a very limited extent) V may designate the manner of C, means of identifying C, or the intended result of C.

4. A SHORT NOTE ON LANGUAGE ACQUISITION

The discussion in the previous sections suggests that an essential cognitive skill for sentence generation is analogy making: that is, *abstraction* and *mapping*. An interesting question is to what extent young children (who acquire their first language) already possess these skills.

Consider, for example, the following example from Berman's (1982) study on the acquisition of Hebrew *binyanim* by children. At the age of two-year-old, Israeli children still fail to use the correct morphological verbal form (suggested to mark underlying mapping configurations, Figure 4). Marked improvement is shown only at the age of three to four year old. The data from Berman suggests however that two and half year old children already master the underlying analogical (abstraction and mapping) operations required for sentence generation, as discussed below. The children only fail to mark the mapping by the correct grammatical form.

In (6) is an example of a sentence generated by Berman's own child around the age of 2;6 (similar examples in English are reported in the CHILDE archive):

(6)	*ima*	*oxelet*	*oti*	*hayom*
	mother	**is-eating**	me	today

(meaning: 'mother is **feeding** me today')

Sentence (6) is syntactically correct (using the simple transitive syntactic construction), with appropriate word order and case marking of nouns. The only error in (6) is that the child used the wrong morphological form for the verb yielding the form *oxelet* ('eating') rather than *ma?axila* ('feeding'). This mistake suggests that the sentence is not a simple imitation of adult's speech (the child has probably never heard the combination 'eat me'), bur rather a real creative production of the child.

The event in the world involves some complex links between the mother and the child (the mother prepares food, then brings it to the child's mouth, thereby enabling the child to eat). But at a higher abstract level, the child correctly perceives this event as analogous to (or an instance of) the basic transitive schema ('Agent affects Patient'), thereby choosing the Transitive syntactic construction to express the event.

The *mapping* performed by the child is also correct. The child perceives the mother as the source (agent) of the causal event and herself as the affected patient, and thus maps the mother to the subject slot and herself to the object slot in the transitive syntactic construction. Into the verbal slot the child maps the *effected* activity of the patient (herself) - 'eating' (rather than, say, the mother's action). This mapping itself is possible in Hebrew (as well as in English, as in *I walked the dog*, where walking refers to the activity of the patient - the dog). The only error the child made is in the morphological marking of the chosen mapping (by *hif'il* morphology, see Figure 4).

To sum, examples such as (6) suggest that a 2.5 year old child already masters the basic cognitive skills (identified in section 2) necessary for sentence generation (*abstraction* and *mapping*). Errors in production at this age may occur only due to lack of command of the *grammatical markers* for these conceptual operations (as suggested for Hebrew morphological *binyanim* above).

5. REFERENCES

Berman, R. (1982). Verb-Pattern Alternation: the Interface of Morphology, Syntax, and Semantics in Hebrew Child Language. *Journal of Child Language*, 9, 161-91.

Damasio, A. (1989). The Brain Binds entities and Events by Multiregional Activation from Convergence Zones. *Neural Computation.*, *1*, 123-132.

Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge: Cambridge University Press.

Fauconnier, G., & Turner, M. (1994). *Conceptual Projection and Middle Spaces*. (Technical Report No. 9401). UCSD: Cognitive Science.

Fauconnier, G., & Turner, M. (1996). Blending as a Central Process of Grammar. In A. Goldberg (Eds.), *Conceptual Structure, Discourse, and Language* Stanford: CSLI.

Fauconnier, G., & Turner, M. (in press). Conceptual Integration Networks. *Cognitive Science.*

Fillmore, C. J., & Kay, P. (1993) *Construction Grammar.* Unpublished manuscript. UCB.

Givón, T. 1984. *Syntax: a Functional-Typological Introduction*, vol 1. Amsterdam: J. Benjamins Pub.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Arguments Structure*. Chicago: Chicago University Press.

Grush, R., & Mandelblit, N. (1998). Blending in Language, Conceptual Structure, and the Cerebral Cortex. In P. Brandt et al. (Eds.), *Acta Linguistica* 31, Copenhagen: Hans Reitzel.

Hofstadter, D. (1995). *Fluid Concepts and Creative Analogies.* Basic Books.

Holyoak, K. & P. Thagard. (1994). *Mental Leaps: Analogy in Creative Thought.* Cambridge: MIT Press.

Idurkhya, B. (1992). *Metaphor and Cognition: An Interactionist Approach.* Boston: Kluwer Academic Publishers.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason.* Chicago: University of Chicago Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago: Chicago University Press.

Lakoff, G. & M. Johnson. (1980). *Metaphors We Live By.* University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar, vol. 1.* Stanford, CA: Stanford University Press.

Langacker, R. W. (1991). *Concept, Image, and Symbol.* New York: Mouton de Gruyter.

Mandelblit, N. (1995). Beyond Lexical Semantics: Mapping of Conceptual and Linguistic Structures in Machine Translation. In *Proceedings of the 4th Int. Conf. on the Cognitive Science of Natural Language Processing*. Dublin, Ireland.

Mandelblit, N. (1997) Grammatical Blending: Creative and Schematic Aspects in Sentence Processing and Translation. Unpublished Ph.D. dissertation. UCSD. *(http://cogsci.ucsd.edu/~mandelbl)*

Mandelblit, N. (ms., submitted). The Grammatical Marking of Conceptual Integration: from Syntax to Morphology.

Mandelblit, N., & Zachar, O. (1998). The Notion of Dynamic Unit: Conceptual Developments in Cognitive Science. *Cognitive Science 22* (2), pp. 229-268.

Mandler, J. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, *99*, 587-604.

Petitot, J. (1995). Morphodynamics and attractor syntax: constituency in visual perception and cognitive grammar, in Port & van Gelder (eds.), *Mind as Motion.* MIT/Bradford.

Shastri, L., & Ajjanagadde, V. (1993). From Simple Associations to Systematic Reasoning: a Connectionist Representation of Rules, Variable, and Dynamic Bindings Using Temporal Synchrony. *Behavioral And Brain Sciences*, *16*(3).

Sweetser, E. (1990). *From etymology to pragmatics: the mind-as-body metaphor in semantic structure and semantic change*. Cambridge: Cambridge University Press.

Talmy, L. (1988). Force Dynamics in Language and Cognition. *Cognitive Science*, *12*, 49-100.

Turner, M. (1996). *The Literary Mind.* Oxford: Oxford University Press.

ANALOGIES IN DESIGN ACTIVITIES
A STUDY OF THE EVOCATION OF INTRA- AND INTERDOMAIN SOURCES

Nathalie Bonnardel & Magali Rech
CREPCO (Research Center in Cognitive Psychology): CNRS UMR 6561 & Université de Provence
29, avenue Robert Schuman 13621 Aix en Provence France
EMail: nathb@newsup.univ-mrs.fr

Analogy-making has been frequently studied in laboratory and on the basis of "well defined" tasks, built towards the end of analyzing specific cognitive mechanisms. Such experiments lead to the proposal of interesting theories and models of analogical reasoning, as for instance the SME model proposed by Gentner (1989) or the approach of Holyak and Thagard (1989). Our objective is in some way different since we wish to study analogy-making on the basis of real-world cognitive activities and, especially, in an area in which analogies can play a very important role: *design activities*.

In non routine design activities, designers have to create an innovative product as well as to satisfy certain specifications. Though certain designers wish to point out the creative and artistic part of their activities (and, for some of them, to keep it in some way "mysterious"), we believe that their creativity can be, at least partially, explained by analogical reasoning, in accordance to certain research works – even not directly related to design – such as the ones of Boden, 1990, Hofstadter 1985, or Kolodner, 1993. Therefore, we settled an experimental situation that should induce non routine design activities as well as allow us to analyze analogy-making by designers and, especially, the effect of classical parameters associated to analogical reasoning (such as intra- vs. interdomain sources). We first characterize more precisely design problem-solving and suggest the role of analogy in it. Then, we describe our experimental situation, present some hypotheses we had as well as the results we obtained. Such results will be finally discussed with regard to certain theoretical approaches of the analogical reasoning.

1. DESIGN PROBLEM-SOLVING AND ANALOGY-MAKING

In Cognitive Psychology, design activities are described as consisting in specific problem-solving, design problems being both *ill defined and open-ended*. Design problems are considered ill-defined because designers have, initially, only an incomplete and imprecise mental representation of the design goals or specifications (Eastman, 1969; Simon, 1973). Design problems are also considered to be open-ended because there is usually no single correct solution for a given problem, but instead a variety of potential solutions (Fustier, 1989). These characteristics lead to design processes involving an iterative dialectic between *problem-framing and problem-solving* (Schoen, 1983; Simon, 1995). During problem-framing, designers refine design goals and specifications and, thus, refine their mental representation of the problem. During problem-solving, designers elaborate solutions and evaluate these solutions with respect to various criteria and constraints (Bonnardel, 1992).

Our general hypothesis is that creativity, which is required for the design of new objects, is dependent on the mental images that the designer can evoke, especially during the problem-framing phase. Such images may be related to objects that are more or less familiar to the designer. More precisely, we believe that these objects can play the role of "sources" (or "bases") for an *analogical reasoning* and, thus, allow the designer to transfer some of the objects' properties to elaborate a target-solution (or target-elements of solution) for the design problem at hand. Though some observations of analogy-making have been made during design activities (see, for instance, Détienne, 1991, and Visser, 1996), we need to analyze more precisely the analogical reasoning in design activities, to understand when designers develop this type of reasoning, how they exploit it and transfer knowledge from one domain to another, etc.

Since creativity in design activities seems to depend on the designers' mental representation, the study we are going to present specifically focus on the evocation part of analogical reasoning, and not on the mapping and transfer parts.

2. EXPERIMENTAL SITUATION

The experimental situation we settled allowed us to control, to some extent, the sources the designers can take into account in order to identify relevant properties for the target and, therefore, construct their own representation of the object to design.

We asked 10 volunteers students in Applied Art (in a technical school of Marseille, France) to design a new product. Though these students are not very experienced designers, they acquired knowledge and skills in design and are really involved in design projects – which, though less complex than those experts deal with, present the main characteristics of professional design projects. Therefore, we will refer to these students in design as "designers".

The design problem they had to solve was defined in collaboration with their professor of Applied Art, in order to have a presentation in accordance with the one used for the design problems they usually have to deal with. Therefore, they were provided with a schedule of conditions consisting, first, in a scenario describing both the object to design and its use (see Figure 1) and, secondly, in a reminder of the main requirements to satisfy.

> The object to be designed was intended to be used in a Parisian "cyber-café". It should be a particular stool with a contemporary design in order to be attractive for young customers. Such stools should allow the user to have a good sitting position, holding the back upright. Towards this end, the users should put their knees on a support intended to this function. In addition, these stools should allow the users to relax, by offering them the possibility to rock.

Figure 1. Brief description of the object to design.

Even for people who are not specialized in design, reading this description involves the evocation of objects we already know. Similarly, the designers can evoke sources to better understand the object to be designed and, eventually, transfer certain properties of the sources to the target. In order to identify the sources evoked by the designers who participated in our study, we asked them to *think aloud* – a method frequently used to study design activities.

The designers' verbalizations as well as their graphical activities were video recorded. Then, the verbalizations were transcribed and matched with the drawing made by the designers.

The experiment was 50 minutes long for each designer. This duration was realistic to realize a rough draft of the object to design. More precisely, it consisted of two phases of 25 minutes each.

1. During the first 25 minutes, two experimental conditions were settled (with 5 designers in each condition):

- a free condition, in which the designers could freely solve the problem and spontaneously evoke sources (known objects) they could refer to;

- a guided condition, in which we presented to the designers names of objects that could play the role of sources. Two of these potential sources for an analogical reasoning were con-

Sources	Intradomain	Interdomain
Studied	"nomadic" stool	logotype
Never studied	rocking-chair	canoe

Table 1. Characteristics of the potential sources proposed to the designers.

sidered *intradomain*, in the sense that they were belonging to the category of "seats". Two other potential sources were considered as *interdomain*, since they refered to objects very different from seats. In addition, one intradomain object and one interdomain object had been studied by the designers during their Art Applied class, whereas the two other objects had never been studied (see Table 1). Each of the names of objects were written on folders and delivered to the designers in a random order.

In this first phase of the experiment, we chose to provide the designers with only *names of objects* – and not graphical representations of specific objects or "instances".

These names refer to categories of objects and may lead the designers to infer what general principle or feature(s) can be extracted from this class of objects as relevant for the object to design. For instance, the designers may reflect on what could be relevant on a canoe or on a logotype for designing the specific stool described in the schedule of conditions.

2. During the following 25 minutes, the designers of the two groups were in a similar situation: they had *both names and a graphical representation* of each type of potential source, what we could call an instance of each category defined by the names (see annex 1).

Designers who belonged to the "guided" group could directly open the folders they had been provided with, to find out the specific graphical representations. During this second phase, designers who belonged to the "free" group were provided with both the names and the graphical representations.

Contrary to the sources' names, their graphical representations facilitate more the identification of relevant principles that the designers can transfer to the object to design. Seeing instances of objects may allow the designers to transfer more directly relevant features to the object to design.

This experimental situation will allow us to determine the influence of potential sources according to the moment of their presentation in the course of the design activity. It will also allow us to compare the influence of the names of objects presented alone with regard to a presentation of both names and instances of sources. However, since we will only analyze the evocation part of analogical reasoning, our analysis will be conducted on "potential" sources for an analogical reasoning. Indeed if some of them effectively lead to a transfer of relevant features to the target, other evoked sources can be more or less rapidly abandoned by the designers.

3. HYPOTHESES

- **Hypothesis 1:**

Our first hypothesis is linked to the progress of the design problem-solving. We expect the role of sources to be more or less important according to the current objectives of the designers. More precisely, the construction by the designers of a mental representation of the object to design can take place more during the beginning of the design problem-solving. Therefore, we expect that *the designers, whatever experimental group they belong to, will evoke less sources as the problem-solving progresses.*

- **Hypothesis 2:**

Our second hypothesis is based on previous research works conducted in Cognitive Psychology and, in particular, on the identification in various domains of a "functional fixation" (see Weisberg, 1988, or, older, Luchins, 1942). For instance, certain studies on the analogical reasoning, conducted with pupils in scholar situation, showed that they tend to systematically reproduce what their teachers showed them as examples (Friemel & Richard, 1988). Such a fixation has also been identified in design activities as "design fixation". Thus, Jansson and Smith (1989, quoted in Purcell & Gero, 1991) showed that the presentation, as examples, of graphical representations of objects that could potentially fit requirements of a design problem, lead designers (and, especially, professional designers) to reproduce numerous features of these objects, comprising features irrelevant to the task at hand.

In accordance to these previous results, in our study, the designers who belong to the guided group could focus on the potential sources we suggested them. Especially, during the first phase, the proposal of names of objects could limit the space of objects that designers can evoke as sources for a design problem-solving based on analogical reasoning. This implies that *the designers who belong to the guided group would evoke less sources than the designers of the free group.* However, we may also observe eventual differences between the presentation of potential sources through names and through instances.

- **Hypothesis 3:**

Though not induced by previous research works, our third hypothesis appears as, partially, in contradiction with the previous one, but allows us to consider more precisely the influence of interdomain sources.

During the first phase, the names of potential sources we presented to the designers who belong to the guided group could, as a *"snowball" effect, lead these designers to consider more sources than the designers of the free group*. The suggestion we made of potential sources a priori independent of the object to design shows to the designers that they can evoke sources that do not belong to the "seat" category and that such a process can present an interesting heuristic power.

4. RESULTS

The previous hypotheses are all based on the number of sources evoked by the designers. Therefore, the results we present are quantitative but they are also related to qualitative features, such as the moment of source evocation with regard to the design problem-solving and the nature of the evoked sources (intra- vs. interdomain). We, now, just present our results and we will comment on them in the section 4.

4.1 Influence of Problem-Solving Phases

The analysis of the evocation of sources by designers was first conducted with regard to the two problem-solving phases we constructed. It showed results in accordance with hypothesis 1:

- The designers evoke a lot more sources during the first 25 minutes than later : a total of 32 evoked sources during the first phase vs. only 8 during the second phase.
- Moreover, it is important to point out that such an effect appears for designers, whatever group they belong to:

  - in the free group, 86% of the sources were evoked during the first phase (which corresponds to, respectively, 6 sources vs. only 1);

  - in the guided group, 79% of the sources were evoked during the first phase (which corresponds to, respectively, 26 sources vs. 7).

4.2 Influence of Experimental Conditions

The analysis of the evocation of sources with regard to the two experimental conditions shows a result opposite to the hypothesis 2:

- The designers who belong to the guided group evoke, in mean, more sources than the designers of the free group: respectively, a total of 33 sources vs. 7, which corresponds in mean to 6.6 sources by designer vs. 1.4 ($p < .05$).
- This effect appears in the two phases of the experiment but is higher in the first phase:

  - during the 1st phase, 26 sources were evoked in the guided condition vs. 6 in the free condition;

  - during the 2nd phase, 7 sources were evoked in the guided condition vs. 1 in the free condition.

4.3. Nature of the Evoked Sources

The analysis of the nature of the sources evoked by the designers of the two group shows results in accordance with the hypothesis 3, about a "snowball effect" of the suggestion of interdomain sources:

The designers who belong to the guided group evoke, in mean, more interdomain sources than the designers of the free group: respectively 3.8 interdomain sources by designer vs. 0.2 ($p < .05$). Therefore, it appears that quite all the sources evoked by the designers of the free group are intradomain whereas the tendency is opposite for the designers of the guided group (see Table 2).

5. DISCUSSION

We comment on our main results with regard to the hypotheses we formulated for this experiment as well as with regard to certain theoretical approaches of the analogical reasoning.

5.1 General Interpretation of the Evocation of Potential Sources

The results we obtained show that designers evoked a lot more sources during the first phase than during the second one. Moreover, we observed that the designers who belonged to the guided group evoked, during the first phase, a lot more sources than the designers of the free group. Such a difference can be due to the "snowball" effect of the potential sources we suggested to the designers. Indeed, we only proposed 4 names of sources, whereas designers of the guided sources evoked 26 sources during the first phase of the experiment. Therefore, it seems that the presentation of names of objects, which refer to categories of these objects, has really a facilitating effect on the evocation process (some interpretations of this fact will be proposed in section 4.2).

Experimental condition Nature of evoked sources	Free condition	Guided condition
Intradomain	6	14
Interdomain	1	19

Table 2. Nature of the evoked sources according to the experimental conditions.

Moreover, again about the design problem-solving phases, we observed that the presentation, for the free group and during the second phase, of the names and instances of potential sources did not have such a facilitating effect. Indeed, though they were presented with such sources, they only evoke 1 source. Therefore the facilitating effect of sources' names appears only at the beginning of design problem-solving. In accordance to this interpretation, the guided group which was provided with instances during this second phase, did not evoke either numerous sources, contrary to what these designers did during the first phase.

To summarize, it seems that the influence of the potential sources we suggested to the designers only appears when they are provided with names of sources and during the first phase of the design problem-solving. Indeed, at the beginning, designers are more involved in the construction of a mental representation of the object to design (i.e., the problem-framing), whereas, later, they are involved in more detailed problem-solving and graphical representation of this object. However, a third experimental condition should have been constructed to decide between the two previous parameters (names and presentation at the beginning) which one has the more important effect: in this last condition, the designers would have been provided directly at the beginning with both names and instances of potential sources. We, initially, planned to have this third experimental condition, but it appeared to be impossible to settle due to the quite limited number of volunteers students who participated in our study. Nevertheless, we can comment more precisely on the influence of the presentation of names of potential sources.

5.2 *Influence of Suggested Sources' Names on Spontaneous Interdomain Sources*

Our second result, about the influence of the names of potential sources for an analogical reasoning, differs from results previously obtained in research areas such as analogical reasoning, problem-solving and design (especially, the results from Friemel & Richard, 1988, and the ones of Jansson & Smith, 1989). Indeed, the presentation of names of potential sources to the designers who belonged to the guided group, did not have an effect of limitation of the space of research of sources that could contribute to solve the design problem through an analogical reasoning. On the contrary, these designers evoked a lot more sources than the designers of the free group. As already expressed, it shows a facilitating effect for the evocation process, which can be explained with regard to two types of interpretations:

1. The effect of "design fixation" may be dependent on the designers' level of expertise: such an effect might become higher as the designers acquire expertise. Experienced designers, such as the professionals who participated in the study of Jansson and Smith (1989), could be more influenced by the suggestion of objects specifically related to the object they have to design (i.e. objects that directly belong to a same category). On the contrary, less experienced designers, such as the students who participated in our study, could be more influenced by objects that are familiar to them, even if these objects are not a priori directly related to the object to design. Other results and, especially, the ones of the study of Purcell and Gero (1991) are also in favor of this interpretation.

Such an interpretation seems to fit particularly design problem-solving. As we pointed out in the characterization of design problems (at the beginning of this text), these problems are open-ended and, thus, allow the designers to refer to various sources. Therefore, less experienced designers or novices have the opportunity to evoke sources that are familiar to them though not directly linked to the object to design (the target).

2. The results we obtained can also be explained by the nature of the sources we suggested to the designers during the 1st phase. These sources are presented as names of objects, by some way related to the object to design. Such names reflect categories of objects and may lead the designers to think of general principles or features that could be transfered to the object to design. Therefore, the designers do not focus on specific features of instances. On the contrary, they can extend their space of research and evoke a diversity of sources, which will have in common with the object to design certain deep principles, for example.

Such an interpretation appears compatible with certain descriptions of the analogical reasoning, proposed on the basis of more traditional experiments. As Ripoll (1998), we can assume the existence of an abstract categorization of objects in long term memory. More precisely, for Ripoll (ibid.), two main types of categories could intervene in the analogical reasoning:

- one, called *"structure tag"* corresponds to the identification of an analogical property category, and is elaborated by the subjects from the structural characteristics of objects – or what we called above deep principles (such as the functioning principle of objects). This structure tag would underly both intra- and interdomain analogical transfers.

- another, called *"domain tag"*, corresponds to the identification of a general semantic category and constitutes a sort of summary of the surface properties of objects. It underlies specifically intradomain analogical transfers.

The third result we obtained allows us to deepen this analysis: the main part of the sources spontaneously[1] evoked by the designers who belonged to the guided group were interdomain sources; whereas the designers who belonged to the free group mainly evoked intradomain sources. Therefore, the facilitating effect on the evocation process seems mainly due to the proposition of interdomain potential sources. For instance, the suggestion of a canoe as potential source shows to the designers that they can be inspired by objects, which a priori seem very far from the object to design. Thus, the role played by the CSTG would become particularly important: the designers would be less focused on surface characteristics of the object they have to design, and they could take into consideration various areas of objects, to look for functioning principles common (at least, partially) to the one they wish to develop for the new object.

---

[1] By "spontaneously" evoked, we mean evoked in addition to the potential sources we suggested.

5.3 Compatible Models of Analogical Reasoning

Some results of this study suggest two main factors that can influence the evocation of sources by designers for an analogical reasoning:

- the *goal of the problem* (i.e., in our study, the object to design).
- the designers' *perception and mental representation of what can constitute potential sources for an analogical reasoning.*

Certain models of the analogical reasoning seem compatible to these suggestions. Especially, we can think of the approach of Holyak and Thagard (1989) takes into account the context and the goal to reach during analogy-making. The importance of the mental representation of the goal of the problem has also been pointed out by Wolstencroft (1989, quoted in Visser, 1989): for this author, the analogical reasoning would be based on a first stage of "identification" that allows an appreciation of the usefulness of mapping. The Copycat model of Mitchell (1989) is also very interesting since it points out the fact that the target and the source have to be perceived as playing the same role at a certain level of abstraction.

CONCLUSION

The analysis we performed was focused on the evocation part of analogical reasoning in design activities. Since such an area of study seems particularly interesting to better understand the creativity developed by designers, we consider that research works towards this end have to be carried on. Concerning our contribution to this perspective, some complementary analyses could be performed on the data we gathered during the experimental situation previously described, in order to determine how the designers use the sources they evoke to solve the problem at hand. Especially, it leads to the study of the evaluation process, which contributes both to the analogical reasoning and to creativity (see Kolodner, 1993). For instance, such a process can be developed to find relevant sources for an analogical reasoning and to determine which particular features of the selected source can be transferred to the target.

In the case of design activities, such studies will contribute to explain how designers can go from the mental representation of known objects to the one of the object to design, until a full and precise graphical representation of the designed object, at the end of design problem-solving.

REFERENCES

Boden, M. (1990). *The Creative Mind: Myths & Mechanisms*. London: Weidenfeld & Nicolson.

Bonnardel, N. (1992). *Le rôle de l'évaluation dans les activités de conception*. Thèse de Doctorat de l'Université de Provence.

Détienne, F. (1991). Reasoning from a schema and from an analog in software code reuse. *Empirical Studies of Programmers: Fourth Workshop*, New Brunswick, N.J., USA, December 6-8.

Eastman, C. M. (1969). Cognitive processes and ill-defined problems: a case study from design. *Proceedings of the First Joint International Conference on I.A.*, Washington, D.C., 669-690.

Friemel, G. & Richard, J.-F. (1988). Apprentissage de l'utilisation d'une calculette. In J.-M. Hoc & P. Mendelsohn, *Les langages informatiques dans l'enseignement. Psychologie française*. Paris: Colin.

Fustier, M. (1989). *La résolution de problèmes : méthodologie de l'action*. Paris : Editions ESF & Librairies Techniques.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds), *Similarity and analogical* *reasoning*, Cambridge: Cambridge University Press, 199-241.

Hofstadter, D.R. (1985). *Metamagical Themas : Questing for the Essence of Mind and Pattern*. New-York: Basic Books.

Holyak, K.J. & Thagard, P.R. (1989). A computational model of analogical problem solving. In S. Vosniadou & A. Ortony (Eds), *Similarity and analogical reasoning*, Cambridge: Cambridge University Press.

Jansson, D.G. & Smith, S.M. (1989). Design fixation. In National Science Foundation, *Proceedings of the Engineering Design Research Conference*, College of Engineering, University of Massachussetts, Amherst, 53-76.

Kolodner, J.L. (1993). Understanding creativity: A case-based approach. In S. Wess, K.-D. Althoff, M.M. Richter (Eds), *Topics in Case-Base Reasoning, Lectures Notes in Artificial Intelligence, № 837*, Berlin: Springer-Verlag, 3-20.

Luchins, A.S. (1942). Mechanization in problem-solving. *Psychological monographs, №*. 248.

Mitchell, M. (1993). *Analogy-Making as Perception: A Computer Model*. Cambridge: The MIT Press.

Purcell, A.T. & Gero, J.S. (1991). The effects of examples on the results of a design activity. In J.S. Gero, *Artificial Intelligence in Design'91*, Oxford: Butterworth-Heinemann Ltd, 525-539.

Ripoll, T. (1998). What this makes me think of that. *Thinking and Reasoning*, 4(1), 15-43.

Schoen, D.A. (1983). *The Reflective Practitioner: How Professionals Think in Action*, New York: Basic Books.

Simon, H.A. (1973). The Structure of Ill Structured Problems. *Artificial Intelligence, № 4*, 181-201.

Simon, H.A. (1995). Problem forming, problem finding and problem solving in design. In A. Collen & W. Gasparski (Eds), *Design & Systems*, New Brunswick (USA): Transaction Publishers, 245-257.

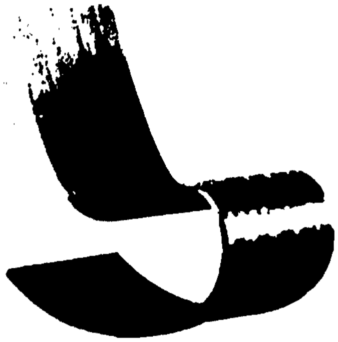
Visser, W. (1996). Two functions of analogical reasoning in design: A cognitive-psychology approach. *Design Studies, № 17*, 417-434.

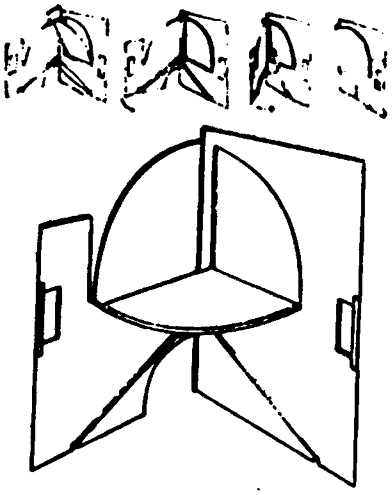
Weisberg, R.W. (1988). Problem solving and creativity. In R.J. Sternberg (Ed.), *The Nature of Creativity: Contemporary Psychological Perspectives*, Cambridge: Cambridge University Press.

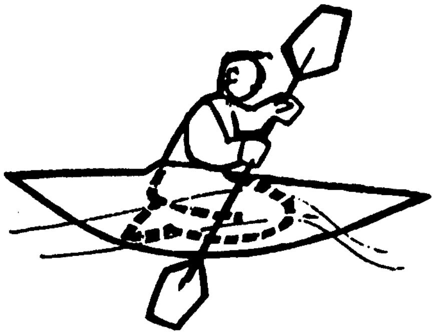
Wolstencroft, J. (1989). Restructuring reminding and repair: What's missing from models of analogy?. *Proceedings of the Scandinavian Conference on A.I.*, Tampere, Finland, June 1989.

Annex 1

Graphical representations of sources proposed to the designers.







'DON'T THINK, BUT LOOK!'
A gestalt interactionist approach to legal thinking

Dan Hunter
Emmanuel College Cambridge CB2 3AP England

Bipin Indurkhya
Tokyo Univ. of Ag. and Tech Nakacho 2-24-16, Koganei. Tokyo 184-8588, Japan

> What is common to [all these games]? — Don't say: "There *must* be something common, or they would not be called 'games' " — but *look and see* whether there is anything common to all. — For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look!
>
> — Wittgenstein, *Philosophical investigations* (emphasis author's)

ABSTRACT

We propose here a new approach to legal thinking that is based on principles of Gestalt perception. Using a Gestalt interaction view of perception, which sees perception as the process of building a conceptual representation of the given stimulus, we articulate legal thinking as the process of building a representation for the given facts of a case. We propose a model in which top-down and bottom-up processes interact together to build arguments (or representations) in legal thinking. We discuss some implications of our approach, especially with respect to modeling precedential reasoning and creativity in legal thinking.

1. INTRODUCTION

We would like to begin by first elaborating on why we use the expression 'legal thinking' and what we mean by it. When talking about what judges, lawyers, law students, and lay people do when applying legal concepts, the convention is to use the expression 'legal reasoning.' We have eschewed the use of this expression however, since it gives the misleading impression that we are talking about inherently rational, indeed logical, thinking. The typical view of law is that it is coherent, internally consistent, logical and rational. Whether or not this true, our interest lies in exposing some of the pre-rational aspects of legal thinking, especially the influence Gestalts have upon the perception of a legal problem. Hence we have not used the term 'legal reasoning' even though at many points — for example in dealing with legal precedent — we will be talking about processes that others would call reasoning.

Having taken this broader view, we would like to note that, on the surface at least, legal thinking and perception seem to have nothing in common. Perception involves receiving some stimulus from the environment, and processing it in some way to integrate it in the conceptual system: it usually involves some kind of identification, representation or description of the stimulus in terms of concepts. Legal thinking, on the other hand, involves generating arguments for a case as to why a certain conclusion follows or does not follow from the given facts of the case: it involves a complex network of rules and statutes, precedents, and several extra-legal factors such as intents of the lawmakers, social and political context and so on. How could these two seemingly different processes be related? Among cognitive processes, legal thinking seems as far removed from perception as one could probably get. How could a model of perception shed any light on generating a legal argument for a given case?

We would like to argue in this paper that, notwithstanding the surface appearances, legal thinking can indeed be viewed as perception. Morover, we would like to show that a certain model of perception, which we refer to as the Gestalt interaction model, can be applied to legal thinking, and in doing so, yields interesting insights into how precedential reasoning works in law, and how its creative aspects can be captured.

This article is organized as follows. In the next section we give provide a sketch of some key ideas and principles that originated from the Gestalt movement. In Section 3, we present some examples of legal thinking that reflect the same principles. In Section 4, we list some key features of our proposed architecture to model legal thinking; and in Section 5 we examine briefly some implications of our proposed view with respect to modeling precendential reasoning and creativity in legal thinking. Finally, Section 6 contains the main conclusions of this article and points to future research issues.

2. GESTALT INTERACTION IN PERCEPTION AND PROBLEM SOLVING

A major finding of the Gestalt school — which was started during the early part of the twentieth century by Duncker, Koffka, Kohler, Luchins, Maier, Wertheimer, and others — was that concepts are more than aggregates of sense data: the human mind prefers to see the world in terms of structured wholes, even when the structure is lacking in the stimuli. The term *Gestalt* was coined to refer to one of these structured wholes. Over the years, the members of this school studied extensively the principles governing Gestalts in perception and problem solving. For example, they articulated two key concepts, namely Einstellung and functional fixity, to explain why some people are unable to solve certain problems, especially in situations where there is a simple, albeit hidden, solution.

Einstellung occurs when a problem solver come to think of certain types of problems as capable of solution in only one way. The best example is Luchins (1942) water jars problem. Subjects were presented with three (usually hypothetical) water jars with varying volumes but no gradations, and asked to measure out a precise goal volume of water. For example, if the volumes of the jars A, B and C are 21, 127, and 3, respectively, and the goal is 100, then a solution of the problem is B-A-2C; meaning that first fill Jar B from a tap, and then from Jar B fill Jar A once (leaving 106 cups in Jar B) and fill Jar C twice (leaving 100 cups). Luchins found that after solving a number of problems where B-A-2C solution applies, subjects fail to see the simpler solution of a problem such as 23, 49, 3, with the goal being 20. For this latter problem, the more complex B-A-2C solution still applies, but a simpler C-A solution is also available. The Einstellung predisposed subjects to solve the water jugs problem in a certain way.

It is worth noting that Einstellung effects can be seen in the representation of a problem, as well as the ability to search the state space of the problem. The water jars experiment is an example of Einstellung in state space search. Kellogg (1995) gives an example of Einstellung in representation. A group of New York mathematics students set their professor the task of finding the next member of the sequence 32, 38, 44, 48, 56, 60. They even hinted that the answer was easy and well-known to the professor. After some complex calculations, the professor generated a difficult mathematical solution. 'No' replied the students, the next member was 'Meadowlark.' They explained that the professor rode the subway everyday: the stops being 32nd St, 38th St, 44th St, 48th St, 56th St, 60th St, and then Meadowlark. Einstellung in representation had meant the professor was unable to see the solution.

Functional fixity is a similar principle to Einstellung, but refers specifically to the use of tools or an object to solve a problem. Studies show that a tool comes to associated with a particular function X, and therefore its use for function Y is often not seen. The quintessential examples are the classic candle problem of Duncker (1945) and two-cord problem of Maier (1931). In the former, subjects were given a candle, a box, drawing pins and a hammer, and asked to fix the candle to a door so that it could be lit. The solution was to hammer the box to the door with the drawing pins, and use it as a stand for the candle. The problem was much more difficult if the box was used to store the candles and drawing pins. The subjects thought of the box's function as that of container only, ignoring its value as a stand. In Maier's experiment, subjects were asked to tie together the free ends of two cords hanging from the ceiling. They were given a number of tools (for example a hammer) but the cords were set further apart than the subject could reach. The solution was to tie the hammer onto one cord, and set it swinging. In this way, the subject could hold one cord, and catch the other one as the newly-created pendulum swung towards them.

Interesting, functional fixity operates in a similar way to Einstellung, in that prior experience can enhance the fixity. So for example Birch and Rabinowitz (1951) had subjects build an electrical circuit prior to the cord problem. The electrical circuit could be completed with either a switch or relay. The subjects were then presented with the cord problem and a prompt. In choosing a pendulum weight they overwhelmingly picked the tool (ie switch or relay) that they had previously not used in the circuit. So 100% of those using the relay in the circuit, used the switch as a weight, and 77% of the switch-users used the relay as a weight. When asked why they had chosen their given tool (ie switch or relay) the subjects explained why it was the only tool available.

These hindrances to problem solving lead to the notions of productive and reproductive thinking (Wertheimer, 1945). Productive thinking involves a recognition of the relations between elements in the problem space (its Gestalt) and the restructuring the elements into a new Gestalt which provides the problem solution. Reproductive thinking is, antithetically, merely the repetition of a learned response.

The difference can be seen in some early work with animals. Thorndike (1911) placed some hungry cats in a box which had a lever in it. The lever opened the door leading to food. The cats would thrash about in the cage and would occasionally knock the level, thereby opening the door. Thorndike showed that having done this a number of times, the cats would gradually learn to hit the lever. This is an example of reproductive thinking. Alternatively there were ape studies of Kohler (1927) where he reported chimpanzees joining two sticks together to reach food outside their cages, in circumstances where they had not been shown how to do this. This type of productive thinking relied on an insight, though it can be improved through hints even where the subject may be unaware of the hint. Maier reported in his two-cord problem that subjects more often reached the pendulum solution when an assistant 'accidentally' brushed against one of the cords setting it in slight motion. And this result occurred even when the subjects could not recall the assistant brushing the cord.This kind of subconscious context effects have also been more recently demonstrated by Kokinov and Yoveva (1996).

Gestalt psychology has enjoyed a recent renaissance, with a number of its findings providing insight into modern research questions (Keane 1988; Garnham and Oakhill 1994). Though the current paradigm in cognitive science focuses on information theory and problem-space conceptions of perception and problem solving, some of the models of the Gestalt school have been re-interpreted in light of information processing theory. (See Brown 1989; Dominowski 1981; Keane 1985, 1989; Newell 1980; Ohlsson 1984a, 1984b, 1985, 1992; Weisberg and Alba 1981, 1982; Weisberg and Suls 1973; and particularly the influential account of vision given by Marr 1982.) Our model of legal thinking as perception follows on in this tradition.

In the information processing model of mind and perception (Lachman, Lachma, and Butterfield, 1979; Eysenck, 1993), information is presented to the organism which is perceived, and then processed, eventually leading to a response. In this model, the starting point is the stimulus from the external environment, which causes certain internal cognitive or conceptual processes. This type of processing is called bottom-up or stimulus-driven processing, since it starts with perception of the most fundamental stimuli at the bottom, and then works its way up into the more abstract conceptual processing system (Eysenck and Keane, 1995; Neisser, 1976). It is involved in most perceptual tasks: understanding the visual field, comprehending phonemes, interpreting touch sensation, and so on. And the main contribution of the Gestalt approach here has been to assert the role of top-down processing in perceptual tasks.

Indeed, while bottom-up processing is clearly important, it is not the whole story. For what you see, depends a great deal on what you want to see, what else is there to see, what else have you seen before, and so on. For example, in spoken word recognition, recognition response times are lower when other lexical, syntactic or semantic information is presented with the word (Marslen-Wilson and Tyler, 1980). Thus, a subject would recognise the word 'butter' more easily if they have just heard the word 'bread' than if they have heard the word 'motor oil' (Eysenck 1993; Tulving, Mandler and Baumel 1964). In Gestalt terms, the prompt of 'bread' will alter the Gestalt we have in the associations between words, and hence alter reaction times to the next word. This is related to the Einstellung findings made by Luchins (1942), described above. In a similar vein, the 'phonemic restoration effect' has also been demonstrated where top-down processing modifies the perception of a single word 'eel' in the following sentences: 'It was found that the eel was on the axle' (wheel), 'It was found that the eel was on the shoe' (heel), 'It was found that the eel was on the orange' (peel) and so forth (Warren and Warren, 1970; Samuel 1981). So we see that both top-down and bottom-up components are two wheels connected to the same axle, and are both necessary for the cognition to proceed. Combining the two approaches we get what we refer here as the Gestalt interaction view.

Though the term 'Gestalt interaction' may be new, the ideas underlying it have been around for quite some time (Neisser, 1976, Pinker, 1985, Ullman, 1985). More recently, one of us (Indurkhya 1992), proposed a formal framework in which concepts and stimuli can interact together to generate 'representations'. Computationally, reasonable models of visual perception and speech recognition have always employed a mix of top-down and bottom-up controls (Erman et al. 1980; Mandal, Murthy & Sankar, 1996; Riseman and Hanson, 1987). It is a similar model that we propose to apply for legal thinking.

3. GESTALT INTERACTION IN LAW

There is a difficulty with the application of gestalt interactionist model of perception to law. That model was developed to explain features of perception: vision processing, word recognition, and the like. Legal reasoning seems to operate at a higher, more abstract level. So we must first identify what, in law, corresponds to stimuli and gestalts, and then proceed to articulate what are the top-down and bottom-up processes.

Generally speaking, legal reasoning starts from the facts of a given case, and proceeds to establish whether certain legal conclusions follow from the facts or not. Whereas the facts of the case are usually expressed in concrete terms, the conclusions involve high-level abstract concepts such as 'negligence', 'duty of care', and so on. Thus, for the first stage in our analysis we can regard the facts as the stimuli, and legal concepts as Gestalts which structure the facts in certain ways.

The implications of this view of legal thinking are fairly obvious. A judge, in deciding a case before her, will be presented with a series of stimuli. These will not be interpreted neutrally. Instead, the existing Gestalt of the judge will dramatically influence her perception of it. Further, as a judge seeks to move from one Gestalt to another, we should be able to see in law Gestalt effects such as Einstellung and functional fixity. Though there is, regrettably, no empirical data on Gestalts in law, we can nonetheless see these effects in one set of data we do have, the legal cases themselves.

Perhaps the most pursuasive demonstration of how Gestalts arise and what a critical role they play in legal thinking, and how they shift over the course of time is made by Levi (1948). In one of his fascinating case study, he showed how the Gestalts 'things imminently dangerous' and 'things inherently dangerous' had a ramarkable influence on the legal issue of liability, and how they have evolved over the last two centuries.

Another good example is the Australian law on whether Aborigines had sovereignty and land rights. Until recently, the indigenous people of Australia had few if any proprietary rights in Australian land. When one considers that the Australian indigenous people had settled the land some 40,000 years prior to the English invasion, this seems unfair. It is even more unfair when one realises that under English law the aborigines should have been granted limited sovereignty over Australia. At the time of the settlement of Australia, English law drew the distinction between lands that were colonised where there was an existing population of people, and lands that were settled where there were no people. Where the land was colonised, the indigenous laws of the people remained, but where the land was empty — or in the Latin *terra nullius* — English law landed at the same moment as the first foot of the British seafarers. Under British colonial rule, Australia was held to be *terra nullius* at the time of white settlement. This was nothing more than a patent fiction, as the evidence of its falsity — the native people, their settlements, their tools, their culture — was present everywhere. Nonetheless the fiction remained and it was held that the only property laws in Australia were those stemming from the introduction of white rule; laws which were less than generous in their grant of land to Aborigines.

The original cases — created during the 1800s in an era of *laissez faire* capitalism and blatant racism — created the initial Gestalt to limit aboriginal holdings of land, except as a consequence of the English property law. Subsequent cases merely adopted the principle that Australia was 'empty land' even though the fiction was always obvious. Each case therefore is a good example of the Einstellung effect, where the perception of the appropriate outcome was set by the previous cases. It is inconceivable that no judge in these cases — whether at trial or during any of the numerous appeals that they entailed — never perceived the term 'empty land' to be at odds with their eventual decision to uphold white rule.

Like the water jar experiments of Luchins (1942) the perception was influenced by Einstellung. However, as with the jars, an alternative Gestalt can supplant the original. This happened in the case of *Mabo v Queensland (No.2).* [(1992) 175 CLR 1]. In *Mabo* the Australian High Court held that previous decisions holding that Australia was *terra nullius* at settlement, and consequently that Aborigines had no indigenous property rights, were wrong at law. This is an interesting decision since the court did not decide to change the law to accommodate modern developments, in the way we see this done in fields as diverse as homicide (including a new defence for 'battered wives') or tax (making modern-day tax evasion illegal) or discrimination law (adding age or sexual-preference as grounds for anti-discrimination suits). Instead the court went back to the basic *terra nullius* formulation at the time of white settlement, and concluded that previous courts were wrong **according to the law at the time**. Notwithstanding prior cases to this effect, the High Court said that Australia could not have been an empty land at settlement, since the Aboriginal presence meant that, at the law of the time, it was a colonised country. Aboriginal law had thus remained in force for the 200 years that the white courts had declared that it never existed. This is a remarkable example of an altered Gestalt, though related processes occur all the time as judges adapt laws to social needs.

Another example is one which focuses on a process that appears to be similar to Maier's two-cord problem and functional fixity. Clearly law does not deal directly with physical tools. However, cases can be seen as one of the tools of legal thinking. This differs somewhat from our earlier characterisation that the case to be decided is a stimulus, but there is no inconsistency here. The Gestalt psychologists realised that perception and problem solving are intimately related, and are both reliant on Gestalts. In the legal field, the Gestalt affects the perception of the current case, as mentioned above. It will also the ability to solve the 'problem' of the case, using the cases available to the judge. These cases then can form their tools, and only some of them are going to be useful to solve any given legal problem. The ability of the judge to use these tools should therefore display similar Gestalt characteristics, including functional fixity. We can demonstrate this with two examples: the first from Anglo-Australian family law and the second from English contract law.

When a married couple divorces, the division of property is determined in large part by the old case law of 'Husband and Wife' and by various Acts. In Australia and England at least, these generally provide for division according to economic added into the marital assets. This was plainly unjust where the husband had worked, while the wife cared for children and maintained the household. In this situation, the standard decision was, until recently, that the husband would get the lion's share of the property. However in an example of productive thinking, one court introduced a principle from a completely different area of law and held that the wife's work placed into the house meant she had an equitable interest in it. The husband, though legally the owner of the house, actually held part of it in 'constructive trust' for his wife. This decision was soon followed by a number of other courts, and is now the standard approach.

This is an example of using a tool — 'constructive trusts' — in a way that was never intended by the original creators of the principle. Another is the decision of Lord Denning in the *High Trees case*, which modified contract law by introducing another equitable principle, this time one called 'promissory estoppel.' The details of this need not detain us, but suffice to say that a legal concept from a different area was drafted into service to deal with a problem in contract law. Both this, and the family law example, show that a type of functional fixity exists in law, but that this can be broken down under pressure.

4. AN ARCHITECTURE FOR LEGAL THINKING

To model legal thinking as Gestalt interaction, we propose an architecture based on 'analogy as high-level perception' approach of Hofstadter and his colleagues (1995), and containing many ideas derived from computational models of perception especially speech recognition (Erman et al. 1982) and machine vision (Riseman and Hanson, 1987; Ullman, 1985). The key features of our proposed architecture are as follows:

- a multi-layer representation is used, with the bottom layer containing the concrete facts, and the top layer containing the Gestalts and the rationale for the decision *(ratio decidendi)* in terms of the Gestalts. Intervening levels contain intermediate concepts and categories that mediate the transition from facts to Gestalts.
- The process of legal thinking is seen as that of coming up with a Gestalt representation in the top layer, given the facts in the bottom layer.
- The process is mediated by both top-down and bottom-up operators. A top-down operator tries to fit the more concrete data of the lower layer into the Gestalt of the upper layer. A bottom-up operator activates a certain Gestalt in the upper layer when a pattern is detected at the lower layer.
- There may also be intra-level operators that connect concepts (Gestalts ot facts) within the same level. They may work in the forward direction (from the conclusions so far reached, derive new conclusions) or in the backward direction (to reach a desired conclusion, posit the necessary sub-conclusions).

- The operators embody statutory knowledge, heuristic knowledge, extra-legal factors, and so on.
- Certain Gestalts may be preactivated in the top layer to reflect the bias or the predisposition of the cognitive agent, or to reflect the current legal doctrines.

5. SOME IMPLICATIONS OF THE PROPOSED VIEW

The model of legal thinking outlined in the last section has some significant implications, especially when compared to the existing approaches to legal reasoning. Here we will briefly examine two such implications.

5.1 Precedential reasoning

The traditional approaches to precedential reasoning in law invariably involve some kind of matching of the facts of the given case with the cases stored in the case library (Ashley, 1990; Branting, 1993). In these approaches, the representations of the cases are kept fixed, so they are not able to model the process of reinterpretation of old cases and Gestalt shifts as new cases are considered, as, for example, recounted in Levi (1948). In our model, however, each case is represented as a multi-layered network connecting the concrete facts of the case with the Gestalts that were found applicable in its decision. And when these networks are activated in order to build a representation for the given facts of a new case, the process is far more complex and subtle than matching parts of the new case against portions of the stored cases. In this process, the old cases are as likely to be reinterpreted as the new case, and it may result in a slight or a drastic change in the Gestalts at the top level.

5.2 Creativity in legal thinking

Though one might expect creativity to be an anathema in legal thinking, we need not look very hard to find many instances where a certain degree of creativity was involved. In such situations, the creativity often lies in the Gestalt switch. In modeling this phenomenon, a key question is: where does the new Gestalt come from? One possible answer to this is that it comes from some other case. One of us has pursued this idea elsewhere (Indurkhya, 1997) to show how creative insights can result from applying a Gestalt from one case to reinterpret another case. In particular, it was shown there how, given two precedents P1 and P2, and a new case N, if P1 and P2 are individually applied to N, a certain conclusion can be derived for the outcome of N; but if the Gestalt of P1 is used to reinterpret P2, and then reinterpreted P2 is applied to N, the opposite conclusion for N can be derived.

6. CONCLUSIONS

We have argued here that Gestalt principles can help us understand a number of features about legal thinking. Notably, it begins to explain why law seems to be a fairly static process of case and rule application. This is due in part to the Einstellung and functional fixity effects inherent in the adoption of one particular Gestalt. It further explains however, why the law goes through upheaval at certain times, as one Gestalt is swapped for another.

This view differs from the traditional, rationalist, formalist view of legal reasoning, where legal concepts are represented as sufficient and necessary conditions, the rigid application of which will lead to perfect justice. This view is one which is rarely accepted these days. Even in Levi's day, it was under attack: "It is important that the mechanism of legal reasoning should not be concealed by its pretense. The pretense is that the law is a system of known rules applied by a judge; the pretense has long been under attack." Levi (1948. p. 1)

Nonetheless, the view that legal reasoning or legal thinking is dependent on formal principles is one that dies hard. In order to advance our Gestalt interactionist model of legal thinking over the formalist view, we need to excavate more carefully what Levi calls the 'mechanism of legal reasoning.' The ideas presented here barely scratch the surface. Just as Gestalt school formulated many principles to explain why certain Gestalts are preferred over others, we also need to articulate in more detail why Gestalts in legal thinking shift the way they do; what necessitates a Gestalt switch; where do the new Gestalts come from; and so on. This would require much empirical work — in terms of case studies and perhaps also experiments involving practising attroneys and judges. From such studies we may be able get a glimpse of what kinds of top-down and bottom-up processes are active in legal thinking, how they are constrained and how they constrain legal Gestalts. We seek to continue this line of research in future, and hope that our ideas will inspire others to join in this endeavour.

REFERENCES

Ashley, K.D. 1990. *Modeling legal arguments: reasoning with cases and hypotheticals.* Cambridge, Mass.: MIT Press.

Birch, H.G. and Rabinowitz, H.S. 1951. The negative effect of previous experience on productive thinking. *Journal of Experimental Psychology 41:* 121-125.

Branting, L.K. 1993. A computational model of ration decidendi. *Artificial intelligence and law 2:* 1-32.

Brown, A.L. 1989. Analogical learning and transfer: What develops? In Vosniadou, S and Ortony, A. (eds) 1989. *Similarity and analogical reasoning,* 369-412, Cambridge: Cambridge University Press.

Dominowski, R.L. 1981. Comment on "an examination of the alleged role of 'fixation' in the solution of several insight problems" by Weisberg & Alba. *Journal of Experimental Psychology: General 110:* 199-203.

Duncker, K. 1945. *On problem-solving,* (trans Lees, L.S.) Psychological Monographs 58 (Whole No. 270) [orig. pub. 1935 in German].

Erman, L.D., Hayes-Roth, F., Lesser, V.R., and Reddy D.R. 1980. The Hearsay-II speech-understanding system: integrating knowledge to resolve uncertainty. *Computing surveys 12:* 213-253.

Eysenck, M.W. 1993. *Principles of cognitive psychology.* Hove: Erlbaum (UK) Taylor & Francis.

Eysenck, M.W. and Keane, M.T. 1995. *Cognitive psychology: A student's handbook.* Hove: Psychology Press

Garnham, A. and Oakhill, J. 1994. *Thinking and reasoning.* Oxford: Blackwell.

Hofstadter, D.H. and the Fluid Analogy Research-Group. 1985. *Fluid concepts and creative analogies.* New York: Basic Books.

Indurkhya, B. 1992. *Metaphor and cognition.* Dordrecht: Kluwer.

Indurkhya, B. 1997. On modeling creativity in legal reasoning. *Proc. of the sixth int. conf. on AI and Law: 180-189.* New York: ACM.

Keane, M.T. 1985. Restructuring revised: A theoretical note on Ohlsson's mechanism of restructuring. *Scandinavian Journal of Psychology, 26:* 363-365.

Keane, M.T. 1988. *Analogical problem solving.* Chicester: Ellis Horwood.

Keane, M.T. 1989. Modelling 'insight' in practical construction problems. *Irish Journal of Psychology, 11:* 201-215.

Kellogg, R.T. 1995. *Cognitive psychology.* Thousand Oaks: Sage.

Kohler, W. 1927. *The mentality of apes.* (2nd ed). New York: Harcourt Brace.

Kokinov, B. and Yoveva M. 1996. Context effects on problem solving. *Proc. of the eighteenth annual conf. of cog. sci. soc.* Hillsdale, NJ: Lawrence Erlbaum.

Lachman, R., Lachman, J.L. and Butterfield, E.C. 1979. *Cognitive psychology and information processing.* Hillsdale: Lawrence Erlbaum.

Levi, E.H. 1948. *An introduction to legal reasoning.* Chicago: University of Chicago Press.

Luchins, A.S. 1942. Mechanisation in problem solving: the effect of Einstellung. *Psychological Monographs 54* (Whole No. 248).

Maier, N.R.F. 1931. Reasoning in humans 2: The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology 12*: 181-194.

Mandal, D.P., Murthy, C.A., and Sankar, K.P. 1996. Analysis of IRS imagery for detecting man-made objects with a multivalued recognition system. *IEEE trans. on systems, man and cybernatics — part A: Systems and humans, Vol. 26:* 241-247.

Marr, D. 1982. *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: Freeman.

Marlsen-Wilson, W.D. and Tyler, L.K. 1980. The temporal structure of spoken language understanding. *Cognition 8:*1-71.

Neisser, U. 1976. *Cognition and reality.* San Francisco: W.H. Freeman.

Newell, A. 1980. Duncker on thinking: An inquiry into progress in cognition. In Koch, S. and Leary, D. (eds), 1980.*A century of psychologyas science: Retrospections and assessments*, New York: McGraw-Hill.

Ohlsson, S. 1984a. Restructuring revisited I: Summary and critique of Gestalt theory of problem solving. *Scandinavian Journal of Psychology, 25:* 65-76.

Ohlsson, S. 1984b. Restructuring revisited II: An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology, 25:* 117-129.

Ohlsson, S. 1985. Retrieval processes in restructuring: Answer to Keane. *Scandinavian Journal of Psychology, 26:* 366-368.

Ohlsson, S. 1992. Information processing explanations of insight and related phenomena. In Keane, M.T. and Gilhooly, K.J. (eds) *Advances in the psychology of thinking.* London: Harvester Wheatsheaf.

Pinker, S. 1985. Visual cognition: an introduction. In S. Pinker (ed.) *Visual cognition,* 1-63. Cambridge, Mass.: MIT Press.

Riseman, E.R. and Hanson A.R. 1987. General knowledge-based vision systems. In M.A. Arbib and A.R. Hanson (eds.) *Vision, brain and cooperative computation,* 285-328. Cambridge, Mass.: MIT Press.

Samuel, A.G. 1981. Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General 110:* 474-494.

Thorndike, E.L. 1911.*Animal intelligence.* New York: Macmillan.

Tulving, E., Mandler, G. and Baumel, R. 1964. Interaction of two sources of information in tachistoscopic word recognition. *Canadian Journal of Psychology 18:* 62-71.

Ullman, S. 1985. Visual routines. In S. Pinker (ed.) *Visual cognition,* 97-159. Cambridge, Mass.: MIT Press.

Warren, R.M and Warren, R.P. 1970. Auditory illusions and confusions. *Scientific American 223:* 30-36.

Weisberg, R.W. and Alba, J.W. 1981. An examination of the alleged role of 'fixation' in the solution of several insight problems. *Journal of Experiment Psychology: General 110:* 169-192.

Weisberg, R.W. and Alba, J.W. 1982. Problem solving is not like perception: More on Gestalt theory. *Journal of Experiment Psychology: General 111:* 326-330.

Weisberg, R.W. and Suls, J. 1973. An information-processing model of Duncker's candle problem. *Cognitive Psychology 4:* 255-276.

Wertheimer, M. 1945. *Productive thinking.* New York: Harper & Row.

Depictive Analogies

Barbara Tversky
Department of Psychology Bldg. 420
Stanford University
Stanford, CA 94305-2130 USA
e-mail: bt@psych.stanford.edu

ABSTRACT

Depictions, such as maps, that portray visible things are ancient whereas depictions, such as graphs and diagrams, that portray things that are inherently not visible, are relatively modern inventions. They serve a variety of functions, such as providing models, attracting attention, supporting memory, facilitating inference and discovery. Depictions use space to convey meaning in ways that are cognitively natural, as suggested by historical and developmental examples. Typically, icons are used to convey elements, based on likenesses and "figures of depiction" and spatial relations are used to convey other relations, based on proximity.

INTRODUCTION

Graphics are one of the oldest and newest form of communication. Long before there was written language, there were pictures, of myriad varieties. A few of the multitude of cave paintings, petroglyphs, bone incisions, clay impressions, stone carvings, and wood markings that people fabricated and used remain from ancient cultures. Some of these prealphabetic depictions probably had religious significance, but many were undoubtedly used to communicate, to keep track of events in time, to note ownership and transactions of ownership, to map places, to record songs and sayings, and to transmit messages (e. g., Coulmas, 1989; De Frances, 1989; Gelb, 1963; Mallery, 1893/1972; Schmandt-Besserat, 1992). As such, they served as permanent records of history, commemorations of cultural past. Because pictures represent meaning more directly than alphabetic written languages, we can guess at their meanings today. In rare cases, we have the benefit of contemporaneous translations. Mallery, for example, was able to speak with native Americans still using pictographic communication as he collected vast numbers of their petroglyphs, birch bark markings (1893/1972).

In many places, the use of pictures to communicate developed into complete written languages. All such languages invented ways to represent concepts that are difficult to depict, such as abstract meanings and proper names. Some pictoric languages transformed and began using written marks to represent the sound of spoken language rather than using marks to represent meaningdirectly. As pictures evolved into written languages, their transparency disappeared. Characters representing abstract concepts were devised and characters representing concrete concepts became schematized and conventionalized. Later, the invention and spread of the alphabet, and then the invention of the printing press decreased reliance on pictures for communication. With the increasing ease of reproducing written language and the spread of literacy, pictures became decorative rather than communicative.

Now, pictures, depictions, and visualizations are on the rise again. As with the proliferation of written language, this is partly due to technologies for reproducing and transmitting pictures. And as with the proliferation of written language, some of the expansion of pictures is due to intellectual insights. For this, the basic insight is using depictions to represent abstract meaning by means of visual and spatial metaphors and figures of depiction. Although depictions have long been used to convey concrete ideas, their use in conveying abstract ideas is more recent. Early depictions for the most part portrayed things that were inherently visualizable, such as objects or environments, in pictographs, maps, or architectural plans. Visualizations of things that are not inherently visualizable, such as temporal, economic, causal, or social relations are a modern invention. These depictions depend on analogy rather than miniaturization or enlargement.

Graphs are perhaps the most prevalent example of depictions of abstract concepts, though not invented until the late eighteenth century (e. g., Beniger and Robyn, 1978; Carswell and Wickens, 1988; Tufte, 1983), although they probably had their roots in mathematical notation, especially Cartesian coordinate systems. Two Europeans, Playfair in England and Lambert in Switzerland, are credited with being the first to promulgate their use, for the most part to portray economic and political data.

Although those early graphs, X-Y plots with time as one of the variables, are still the most common type of graph in scientific journals (Cleveland, 1984), varieties of graphs, graphics, and visualizations abound, with new ones appearing all the time. Bar graphs and pie charts are common for representing quantitative data, with flow charts, trees, and networks widely used for qualitative data. Icons appear in airports, train stations, and highways all over the world, and menus of icons on information highways over the world. Many are used to portray concepts that are difficult to visualize.

The choices of icons and graphic displays are usually not accidental or arbitrary. Many have been invented and reinvented by adults and children across cultures and time. Many have analogs in language and in gesture and parallels in Gestalt principles of perceptual organization. They seem rooted in natural cognitive correspondences, "figures of depictions," and spatial metaphors.

In this paper, I present an analysis of graphic displays based on their functions and on their structure. The evidence I will bring to bear is eclectic and unconventional, drawing from examinations of historical graphic inventions, children's graphic inventions, and language.

Other Approaches. Others have taken a broad view of graphics from other perspectives. Bertin (1981) put forth a comprehensive semiotic analysis of the functions of graphics and the processes used to interpret them that established the field and defined the issues. According to Bertin, the functions of graphs are to record, communicate, and process information, and the goal of a good graphic is simplification to those ends. Ittelson (1996) has pointed to differences in processing of "markings," deliberate, two-dimensional inscriptions on surfaces of objects and other visual stimuli. Winn (1987) has discussed how information is conveyed in charts, diagrams, and graphs. Larkin and Simon (1987) have examined the differences between sentential and diagrammatic external representations, pointing to the advantages of diagrammatic ones for tasks where spatial proximity conveys useful information. Stenning and Oberlander (1995) have analyzed the advantages and disadvantages of diagrammatic and sentential representations in drawing inferences. They argue that diagrams allow expression of some abstractions, much like natural language, but are not as expressive as sentential logics. Cleveland (1984; 1985) has examined the psychophysical advantages and disadvantages of using different graphic elements, position, angle, length, slope, and more, for efficiency in extracting different kinds of information from displays of quantitative data. He and his collaborators have produced convincing cases where conventional data displays can be easily misconstrued by human users. Tufte (1983, 1990, 1997) has exhorted graphic designers to refrain from "chart junk," extraneous marks that convey no additional information, adopting by contrast a minimalist view. Wainer (1984, 1992) has gathered a set of useful prescriptions and insightful examples for graph construction, drawing on work in semiotics, design, and information processing. Kosslyn (1989; 1994), using principles adopted from visual information processing and Goodman's (1978) analysis of symbol systems, has developed a set of prescriptives for graphic design, based on an analysis of the syntax, semantics, and pragmatics underlying graphs. Pinker (1990) provides an analysis of information extraction from graphics that separates processes involved in constructing a visual description of the physical aspects of the graph from those involved in constructing a graph schema of the mapping of the physical aspects to mathematical scales. Carswell and Wickens (Carswell, 1992; Carswell & Wickens, 1988; 1990) have demonstrated effects of perceptual analysis of integrality on graph comprehension, and others have shown biases in interpretation or memory dependent on graphic displays (Gattis & Holyoak, 1996; Levy, Zacks, Tversky, & Schiano, 1996; Schiano & Tversky, 1992; Shah & Carpenter, 1995; Spence & Lewandowsky, 1991; Tversky & Schiano, 1989).

SOME FUNCTIONS OF GRAPHIC DISPLAYS

Despite their variability of form and context, a number of cognitive principles underlie graphic displays. These are evident in the many functions they serve as well as in the way information is conveyed in them. Some of their many overlapping and sometimes conflicting functions are sketched below. As with functions, goals, and constraints on other aspects of human behavior, so the functions of graphic displays are sometimes at odds with each other.

Attract attention and interest. One prevalent function of graphic displays is to attract attention and interest. As such, graphics may be pleasing or shocking or repulsive or calming or funny.

Models of actual and theoretical worlds. Maps, architectural drawings, molecules, circuit diagrams, organizational charts, flow diagrams are just some of the myriad examples of diagrams serving as models of worlds and the things in them. Note that these are models, and not strictly shrunken or expanded worlds. Effective diagrams omit features that are in the modeled world, distort others, and add features that are not in the modeled world. Maps, for example, may exaggerate the sizes of streets so that they can be seen. They introduce symbolic elements, for railroads, ocean depth, towns, and more, that require a key and/or convention to interpret. The essence of creating an effective externalrepresentation is to abstract those features that are essential and to eliminate those that are not.

Record information. An ancient function of graphics is preserving records. Tallies, for example, were devised to keep track of property, beginning with a simple one mark for one item relation, developing into numerals as tallies became cumbersome for large sums and calculations (Schmandt-Besserat, 1992).

Facilitate memory. Facilitating memory was surely was and is one of the functions of writing, whether pictographic or alphabetic. A contemporary example is the use of computer menus, which turn a recall task into a recognition one. Graphical user interfaces promote memory in another way, by using spatial locations cues, an ancient device, the Method of Loci, with modern support (e. g., Bower, 1970; Franklin, Tversky, and Coon, 1992; Small, 1997; Taylor and Tversky, 1997; Yates, 1969).

Communication. In addition to facilitating memory, graphic displays also facilitate communication. As for memory, this has also been an important function of writing, to allow communication out of earshot (or eyeshot). Graphic displays allow private, mental conceptualizations to be made public, where they can be shared, examined, and revised.

Effective graphics make it easy for users to extract information and draw inferences from them. Maps, for example, facilitate determining routes and estimating distances. A map of cholera cases in London during an epidemic made it easier to find the contaminated water pump (Wainer, 1992). Plotting change rather than absolute levels of a measure can lead to very different inferences (Cleveland, 1985). Indeed, the advice in How to Lie with Statistics (Huff, 1954) has been used for good or bad over and over. Physics diagrams (Narayanan, Suwa, & Motoda, 1994) and architectural sketches (Suwa &

Tversky, 1996) bias users towards some kinds of inferences more readily than others.

Graphic displays accomplish all these functions and more in two separable ways, through the use of graphic elements or icons, and through the spatial array of elements. Different cognitive principles underlie each. In general, graphic elements are used to represent elements in the world and graphic space is used to represent the relations between elements, though there are exceptions to this generalization. This dichotomy into elements and relations maps loosely onto the "what" vs. "where" distinction in vision and in spatial cognition.

The fact that graphic displays are external representation devices augments many of their functions. Spatially organized information can be accessed and integrated quickly and easily, especially when the spatial organization reflects conceptual organization. Several people can simultaneously inspect the same graphic display, and refer to it by pointing and other devices in ways apparent to all, facilitating group communication.

ICONS: FIGURES OF DEPICTION

Sometimes icons can be used to represent meaning directly, for example, highway signs portraying a picnic table or falling rocks to indicate the presence of actual ones. "Figures of depiction," analogous to figures of speech, can be used to portray concepts that are not readily depicted (Tversky, 1995). One common type of figure of depiction is metonymy, where an associated object represents the concept. Returning to computer interfaces, a picture of a folder can represent a file of words and a picture of a trash can represent a place for unwanted folders. Analogous examples in language include using "the crown" to represent the king and "the White House" to represent the president. Synecdoche, where a part is used to represent a whole, or a whole for a part, is another common figure of depiction. In highway signs, an icon of a place setting near a freeway exit indicates a nearby restaurant and an icon of a gas pump a nearby gas station. Analogous examples in language include "give a hand" for help and "head count" for number of people. These same figures of depiction are frequent in icons in early pictographic writing (Coulmas, 1989; Gelb, 1963; Tversky, 1995). For example, early Sumerian writing used a foot to indicate "to go" and an ox's head to indicate an ox. Children's spontaneous writing and depictions also illustrate these principles (e. g., Hughes, 1986; Levin and Tolchinsky-Landsman, 1989). Like the inventors of pictographic languages, children find it easier to depict objects, especially concrete ones, than operations. For abstract objects and operations, children use metonymy and synecdoche. For example, children draw hands or legs to indicate addition or subtraction. Interestingly, the latter was also used in hieroglyphics.

The meanings of these depictions are somewhat transparent. Often, they can be guessed, sometimes with help of context, and even when guessing is not likely, they are easily associated to their meanings, and thus easily remembered. (for similar arguments in the context of ASL and gesture, see Macken, Perry and Haas, 1993). Depictions have other advantages over words. Meaning is extracted from pictures faster than from words (Smith and McGee, 1980). Icons can be "read" by people who do not read the local language.

A new use of depictions has appeared in email, emotions. Seemingly inspired by smiley faces, and probably because it is inherently more casual than other written communication, computer vernacular has added signs for the emotional expression normally conveyed in face-to-face communication by intonation and gesture. These signs combine symbols found on keyboards to denote facial expressions, usually turned 90 degrees, such as :) or ;).

GRAPHIC ARRAYS: SPATIAL METAPHORS

Graphs, charts, and diagrams convey qualitative and quantitative information using natural correspondences and spatial metaphors. The most basic of the metaphors is proximity: proximity in space is used to indicate proximity on some other property, such as time or value. Spatial arrays convey conceptual information metaphorically at different levels of precision, corresponding to the four traditional scale types, nominal, ordinal, interval, and ratio (Stevens, 1946). These are ordered inclusively by the degree of information preserved in the mapping. Spontaneously produced graphic displays reflect these scale types. Children, for example, represent nominal relations in graphic displays at an earlier age than ordinal relations, and ordinal relations at an earlier age than interval relations (Tversky, Kugelmass, and Winter, 1991).

Nominal scales are essentially clusters of elements sharing a single property or set of properties. Graphic devices indicating nominal relations often use the simplest form of proximity, grouping (cf. Gestalt Principle of Grouping). Things that are related are placed contiguously or in close proximity, spatially separated from unrelated things. One use of this device that we take for granted is the spaces between words in writing. In early writing, there were no spaces between words. Another example of using separation in space to indicate separation of ideas is indentation and/or spacing for paragraphs.

A list provides another spatial device for delineating a category, where all the items that need to be purchased or tasks that need to be done are written in a single column. Items are separated by empty space, and the items begin at the same point in each row, indicating equivalence. For lists, there is often only a single category; organization into a column indicates that the items are not randomly selected, but rather, share a property. Multiple lists are also common, for example, the list of chores of each housemate. A table is an elaboration of a list, using the same spatial device to organize both rows and columns (Stenning and Oberlander, 1995). Examples include a list of countries with their GNP's for each of the last ten years, or a list of schools, with their average achievement scores on a variety of tests. Tables cross-classify. Items within each column and within each row are related, but on different features. For example, columns may correspond to countries and GNP's by year, or to schools and scores by test, and rows may provide the values for each country or school. Train schedules are yet another example, where the first column is typically the stops and subsequent columns are the times for each train. For train schedules, a blank space where there would ordinarily be a time indicates a non-event, that is, this train doesn't stop at that station. Using spatially-arrayed rows and columns, tables group and juxtapose simultaneously.

Special signs, usually visual ones rather than strictly spatial ones, are sometimes used to indicate grouping. These seem to fall into two classes, those based on linking or enclosure (cf. Gestalt Principle of Grouping) and those based on similarity (cf. Gestalt Principle of Similarity). Many signs used for grouping resemble physical structures that enclose things, such as bowls and fences, or physical structures that link things, such as paths. Some analogous structures on paper are lines, parentheses, circles, boxes, and frames. Like paths or outstretched arms, lines link one concept to another, bringing noncontiguous things into contiguity, making distal items proximal. In tables, lines, sometimes whole (_____), sometimes partial (.......) (one might interpret broken lines as more tentative than solid ones), are used to link related items. Tables often add boxes to emphasize the structures of rows and columns or to enclose related items and separate different ones. Newspapers use boxes to distinguish one classified ad from another. Parentheses and brackets in writing are in essence degenerate circles. The curved or bent lines, segments of circles or rectangles, face each other to enclose the related words and to separate them from the rest of the sentence.

Circles indicating items belonging to the same set are useful in visualizing syllogisms and in promoting inference as in Euler or Venn diagrams or in contemporary extensions of them (e. g., Shin, 1991; Stenning and Oberlander, 1995). Circles with no physical contact indicate sets with no common items, and physically overlapping circles indicate sets with at least some common items. To increase the inferential power of Euler diagrams, spatial signs based on similarity have been added, such as filling in similar regions with similar and dissimilar regions with different marks, color, shading, cross-hatching, and other patterns (e. g., Shin, 1991). Maps use colors as well as lines to indicate political boundaries and geographic features. For geographic features, many of the correspondences are natural ones. For example, deserts are colored beige whereas forests are colored green, and lakes and seas are colored blue, with darker (deeper) blues indicating deeper water.

Ordinal relations can vary from a partial order, where one or more elements have precedence over others, to a complete order, where all elements are ordered with respect to some property or properties. There are two separable issues in mapping order onto space. One is the devices used to indicate order, and the other is the direction of order. They will be discussed in order. Writing is ordered, so one of the simplest spatial devices to indicate rank on some property is to write items according to the order on the property, for example, writing countries in order of GNP, or people in order of age. Degrees of empty space can be used to convey order, as in progressive indentation in outlines.

Lines can be used to indicate order as well as equivalence. Lines form the skeletons of trees and graphs, both of which are commonly used to display ordered concepts, to indicate asymmetry on a variety of relations, including kind of, part of, subservient to, and derived from. Examples include hierarchical displays, as in linguistic trees, evolutionary trees, and organizational charts. Other visual and spatial devices used to display order rest on the metaphor of salience. More salient elements have more of the relevant property, be it size, , color, highlighting, or superposition. Some of these devices rely on what can be called natural cognitive correspondences. For example, high temperatures are associated with "warm" colors and low temperatures with "cold" colors, as used in weather maps and scientific charts. This association most likely derives from the colors of things varying in temperature, such as fire and ice.

Arrows are a special kind of line, with one end marked, inducing an asymmetry. Although they have many uses, a primary one is to indicate direction, an asymmetric relation. Arrows seem to be based on either or both of two spatial analogs. One obvious analog is the projectile, invented by many different cultures for hunting. It is not the hunting or piercing aspects of physical arrows that have been adopted in diagrams, but rather the directionality. Hunting arrows are asymmetric as a consequence of which they fly more easily in one direction than the other. Another analog is the idea of convergence captured by the > ("V") of a diagram arrow. Like a funnel or river straits, it directs anything captured by the wide part to the point, and straight outwards from there. Arrows are frequently used to signal direction in space. In diagrams, arrows are also commonly used to indicate direction in time. In production charts and computer flow diagrams, for examples, arrows are used to denote the sequence of processes. Terms for time, such as "before" and "after," and indeed thinking about time, frequently derive from terms for and thinking about space (e. g., Clark, 1973).

Interval and ratio relations apply more constraints of the spatial proximity metaphor than ordinal relations. In graphic displays of interval information, the spaces between elements are meaningful; that is, greater space corresponds to more on the relevant dimension. This is not the case for ordinal mappings. In displays of ratio information, the ratios of the spaces are meaningful.

The most common graphic displays of interval and ratio information are X-Y plots, where distance in the display corresponds to distance on the relevant property or properties. Bar charts are useful for displaying quantities for several variables at once; here, the height or length of the bar corresponds to the quantity on the relevant variable. Isotypes combine icons and bar charts to render quantities on different variables more readily interpretable (Neurath, 1936). For example, in order to display the yearly productivity by sector for a number of countries, a unit of output for each sector is represented by an isotype, or icon that is readily interpretable, a shaft of wheat for grain, an ingot for steel, an oil well for petroleum. The number of icons per sector is proportional to output in that sector. Icons facilitate comparison across countries or years for the same sector. Isotypes were invented by Otto and Marie Neurath in the 30's as part of a larger movement to increase communication across languages and cultures. That movement included efforts to develop picture languages and Esperanto. Musical notation is a specialized interval scale that makes use of a limited visual alphabet corresponding to modes of execution of notes as well as a spatial scale corresponding to pitch. Finally, for displaying ratio information, pie charts can be useful, where the area of the pie corresponds to the proportion on the relevant variable.

DIRECTIONALITY

In spite of the uncountable number of possibilities for indicating order in graphic displays, the actual choices are remarkably limited. In principle, elements could be ordered in any number of orientations in a display. Nevertheless, graphic displays tend to order elements either vertically or horizontally or both. Similarly, languages are written either horizontally or vertically, in rows or in columns. There are reasons grounded in perception for the preference for vertical and horizontal orientations. The perceptual world has a vertical axis defined by gravity and by all the things on earth correlated with gravity and a horizontal axis defined by the horizon and by all the things on earth parallel to it. Vision is especially acute along the vertical and horizontal axes (Howard, 1982). Memory is poorer for the orientation of oblique lines, and slightly oblique lines are perceived and remembered as more vertical or horizontal than they were (Howard, 1982; Schiano and Tversky, 1992).

Of all the possible orientations, then, graphic displays ordinarily only use the vertical and horizontal. What's more, they use these orientations differently. Vertical arrays take precedence over horizontal ones. Just as for the choice of dimensions, the precedence of the vertical is also rooted in perception (Clark, 1973; Cooper and Ross, 1975; Lakoff and Johnson, 1980; Franklin and Tversky, 1990). Gravity is correlated with vertical, and people are oriented vertically. The vertical axis of the world has a natural asymmetry, the ground and the sky, whereas the horizontal axis of the world does not. The dominance of the vertical over the horizontal is reflected in the dominance of columns over rows. It is more usual and more natural to make a vertical list than a horizontal one. Similarly, bar charts typically contain vertical columns.

There is another plausible reason for the dominance of the vertical over the horizontal. Not only does the vertical take precedence over the horizontal, but there is a natural direction of correspondence for the vertical, though not for the horizontal. In language, concepts like more and better and stronger are associated with upward direction, and concepts like less and worse and weaker with downward direction (Clark, 1973; Cooper and Ross, 1975; Lakoff and Johnson, 1980). People and plants, indeed most life forms, grow upwards as they mature, becoming bigger, stronger, and (arguably) better. Healthy and happy people stand tall; sick or sad ones droop or lie down. More of any quantity makes a higher pile. The associations of up with quantity, mood, health, power, status, and more derive from physical correspondences in the world. It is no accident that in most bar charts and X-Y plots, increases go from down to up. The association of all good things with up is widely reflected in language as well (inflation and unemployment are exceptions, but principled ones, as the numbers used to convey inflation and unemployment go up). We speak of someone "at the top of the heap," of doing the "highest good," of "feeling up," of being "on top of things," of having "high status" or "high ideals," of doing a "top-notch job," of reaching "peak performance," of going "above and beyond the line of duty." In gesture, we show success or approval with thumbs up, or give someone a congratulatory high five. The correspondence of pitch with the vertical seems to rest on another natural cognitive correspondence. We produce higher notes at higher places in the throat, and lower notes at lower places. It just so happens that higher notes correspond to higher frequency waves, but that may simply be a happy coincidence.

In contrast, the horizontal axis is standardly used for neutral dimensions, for example, time. Similarly, with the major exception of economics, neutral or independent variables are plotted along the horizontal axis, and the variables of interest, the dependent variables, along the vertical axis. Although graphic conventions stipulate that increases plotted horizontally proceed from left to right, directionality along the horizontal axis does not seem to rest in natural correspondences. The world is asymmetric along the vertical axis, but not along the horizontal axis. Right-left reflections of pictures are hardly noticed but top-bottom reflections are (e. g., Yin, 1969). Languages are just as likely to be written left to write as right to left (and in some cases, both), but they always begin at the top. Children and adults from cultures where language is written left to right as well as from cultures where language is written right to left mapped increases on a variety of quantitative variables from down to up, but almost never mapped increases from up to down. However, people from both writing cultures mapped increases in quantity and preference from both left to right and right to left equally often. The relative frequency of using each direction to represent quantitative variables did not depend on the direction of written language (Tversky, et al, 1991). Despite the fact that most people are right-handed and that terms like dexterity derived from "right" in many languages have positive connotations and terms like sinister derived from "left" have negative connotations, the horizontal axis in graphic displays seems to be neutral. Consistent with that, we refer to one side of an issue as "on the one hand," and the other side as "on the other hand," which has prompted some politicians to ask for one-handed advisors. And in politics, both the right and the left claim the moral high ground.

Children's and adults' mappings of temporal concepts showed a different pattern from their mappings of quantitative and preference concepts (Tversky, et al, 1991). For time, they not only preferred to use the horizontal axis, they also used the direction of writing to determine the direction of temporal increases, so that people who wrote from left to right tended to map temporal concepts from left to right and people who wrote from right to left tended to map temporal concepts from right to left. This pattern of findings fits with the claim that neutral concepts such as time tend to be mapped onto the horizontal axis. The fact that the direction of mapping time corresponded to the direction of writing but the direction of mapping quantitative variables did not may be because temporal sequences seem to be incorporated into writing more than quantitative concepts, for example, in schedules, calendars, invitations, and announcements of meetings.

Consistent with the previous arguments and evidence, ordinal charts and networks tend to be vertically organized. A survey of the standard scientific charts in all the textbooks in biology, geology, and linguistics at the Stanford Undergraduate Library revealed vertical organization in all but two of 48 charts (Tversky, 1995). Furthermore, within each type of chart, there was agreement as to what appeared at the top. In 17 out of the 18 evolutionary charts, Homo sapiens, that is, the present age, was at the top. In 15 out of the 16 geological charts, the present era was at the top, and in 13 out of the 14 linguistic trees, the proto-language was at the top. In these charts, in contrast to X-Y graphs, time runs vertically, but time does not seem to account for the direction, partly because time is not ordered consistently across the charts. Rather, at the top of each chart is an ideal. In the case of evolution, it is humankind, regarded by some as the pinnacle of evolution, a view some biologists discourage. In the case of geology, the top is the richness and accessibility of the present era. In the case of language trees, the top is the protolanguage, the most ancient theoretical case, the origin from which others diverged. In organizational charts, say of the government or large corporations, power and control are at the top. For diagramming sentences or the human body, the whole is at the top, and parts and sub-parts occupy lower levels. In charts such as these, the vertical relations are meaningful, denoting an asymmetry on the mapped relation, but the horizontal relations are often arbitrary.

BASIS FOR METAPHORS AND COGNITIVE CORRESPONDENCES

A major purpose of graphic displays is to represent visually concepts and relations that are not inherently visual. Graphic displays use representations of elements, primarily icons, and the spatial relations among them to do so. To enhance communication, both elements and relations are based on people's perception of and interaction with the familiar physical world, especially the spatial world. People have extensive experience observing and interacting with the physical world, and consequently extensive knowledge about the appearance and behavior of things in it. It is natural for this concrete experience and knowledge to serve as a basis for pictorial, verbal, and gestural expression.

Naturalness is found in natural correspondences, "figures of depiction," and spatial metaphors, derived from extensive human experience with the concrete world. It is revealed in language and in gesture as well as in a long history of depictions.

REFERENCES

Beniger, J. R. & Robyn, D. L. (1978). Quantitative graphics in statistics. *The American Statistician, 32*, 1-11.

Bertin, J. (1981). *Graphics and graphic-information-processing*. N. Y.: Walter de Gruyter.

Bower, G. H. (1970). Analysis of a mnemonic device. *American Scientist, 58*, 496-510.

Carswell, C. M. (1992). Reading graphs: Interaction of processing requirements and stimulus structure. In B. Burns (Ed.), *Percepts, concepts, and categories*. Pp. 605-645. Amsterdam: Elsevier.

Carswell, C. M. and Wickens, C. D. (1988). Comparative graphics: history and applications of perceptual integrality theory and the proximity compatibility hypothesis. Technical Report, Institute of Aviation, University of Illinois at Urbana-Champaign.

Carswell, C. M. & Wickens, C. D. (1990). The perceptual interaction of graphic attributes: Configurality, stimulus homogeneity, and object integration. *Perception and Psychophysics, 47*, 157-168.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. Pp. 27-63. New York: Academic Press.

Cleveland, W. S. (1984). Graphs in scientific publications. *The American Statistician, 38*, 261-269.

Cleveland, W. S. (1985). *The elements of graphing data*. Monterey, CA: Wadsworth.

Coulmas, F. (1989). *The writing systems of the world*. Oxford: Basil Blackwell.

Cooper, W. E. & Ross, J. R. (1975). World order. In R. E. Grossman, L. J. San, & T. J. Vance, (Eds.), *Papers from the Parasession on Functionalism*. Chicago: Chicago Linguistic Society.

DeFrances, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.

Franklin, N. and Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General, 119*, 63-76.

Franklin, N., Tversky, B., and Coon, V. (1992). Switching points of view in spatial mental models acquired from text. *Memory and Cognition, 20,* 507-518.

Gattis, M. and Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1-9.

Gelb, I. (1963). *A study of writing.* Second edition. Chicago: University of Chicago Press.

Goodman, Nelson. *Languages of art: An approach to a theory of symbols.* New York: Bobbs-Merrill, 1968.

Howard, I. P. (1982). *Human visual orientation.* New York: Wiley.

Huff, D. (1954). *How to lie with statistics.* New York: Norton.

Hughes, M. (1986*). Children and number: Difficulties in learning mathematics.* Oxford: Blackwell.

Ittelson, W. H. (1996). Visual perception of markings. *Psychonomic Bulletin & Review, 3,* 171-187.

Kosslyn, S. M. (1989) Understanding Charts and Graphs. *Applied Cognitive Psychology, 3,* 185-223;

Kosslyn, S. M. (1994). *Elements of graph design.* New York: Freeman.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Larkin, J. H. and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11,* 65-99.

Levin, I. & Tolchinsky Landsmann, L. (1989). Becoming literate: Referential and phonetic strategies in early reading and writing. *International Journal of Behavioural Development, 12,* 369-384

Levy, E., Zacks, J., Tversky, B. and Schiano, D. (1996). Gratuitous graphics: Putting preferences in perspective. *Human factors in computing systems: Conference proceedings* (pp. 42-49). NY: ACM.

Macken, E., Perry, J. and Haas, C. (1993). Richly grounded symbols in ASL. *Sign Language Studies, 81,* 375-394.

Mallery, G. (1893/1972). *Picture writing of the American Indians.* (Originally published by Government Printing Office). NY: Dover.

Narayanan, N. H., Suwa, M., and Motoda, H. (1994). A study of diagrammatic reasoning from verbal and gestural data. *Proceedings of the 16th Annual Conference of the Cognitive Science Society.*

Neurath, O. (1936). *International Picture Language: The First Rules of Isotype.* London: Kegan Paul, Trench, Trubner & Co., Ltd.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.*), Artificial intelligence and the future of testing.* Pp. 73-126. Hillsdale, N. J.: Erlbaum.

Schiano, D. and Tversky, B. (1992). Structure and strategy in viewing simple graphs. *Memory and Cognition, 20,* 12-20.

Schmandt-Besserat, D. (1992). Before writing, Volume 1: From counting to cuneiform. Austin: University of Texas Press.

Shah, P. and Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General, 124,* 43-61.

Small, J. P. (1997) *Wax tablets of the mind.* New York: Routledge, Paul.

Smith, M. C. and McGee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General, 109,* 373-392.

Spence, I. & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology, 5,* 61-77.

Stenning, K. and Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science, 19,* 97-140.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 677-680.

Suwa, M. & Tversky, B. (1996). What architects see in their sketches: Implications for design tools. *Human factors in computing systems: Conference companion* (pp. 191-192). NY: ACM.

Taylor, H. A. and Tversky, B. (1997). Indexing events in memory: Evidence for index preferences. *Memory, 5,* 509-542.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Tufte, E. R. (1990). *Envisioning information.* Cheshire, CT: Graphics Press.

Tufte, E. R. (1997). *Visual explanations.* Cheshire, CT: Graphics Press.

Tversky, B. (1995). Cognitive origins of graphic conventions. In F. T. Marchese (Editor). *Understanding images.* Pp. 29-53. New York: Springer-Verlag.

Tversky, B., Kugelmass, S. and Winter, A. (1991) Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23,* 515-557.

Tversky, B. and Schiano, D. (1989). Perceptual and conceptual factors in distortions in memory for maps and graphs. *Journal of Experimental Psychology: General, 118,* 387-398.

Wainer, H. (1980). Making newspaper graphs fit to print. In P. A. Kolers, M. E. Wrolstad and H. Bouma (Editors), *Processing of visible language 2.* Pp. 125-142. NY: Plenum.

Wainer, H. (1984). How to display data badly. *The American Statistician, 38,* 137-147.

Wainer, H. (1992) Understanding graphs and tables. *Educational Researcher, 21,* 14-23.

Winn, W. D. (1987). Charts, graphs and diagrams in educational materials. In D. M. Willows and H. A. Haughton (Eds.). *The Psychology of illustration.* N. Y.: Springer-Verlag.

Yates, F. A. (1969). *The art of memory.* New York: Penguin.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141-45.

ANALOGY AND INDUCTION : WHICH (MISSING) LINK ?

Antoine Cornuéjols & Jacques Ales-Bianchetti
Laboratoire de Recherche en Informatique (LRI), UA 410 du CNRS
Université de Paris-sud, Orsay
Bâtiment 490, 91405 ORSAY (France)
email : {antoine,ales}@lri.fr

ABSTRACT

In this paper, we argue that accounts of analogy should be consistent with the theoretical frameworks developed for related cognitive processes, such as induction. On one hand, this allows to more firmly anchor our theoretical perspectives on analogy, and, on the other hand, this may offer ways to improve on the current theories in the related fields. We propose some steps towards these goals.

1. INTRODUCTION

The study of analogy confronts us with a formidable challenge. Its manifestations are seemingly ubiquitous : from perceptual processes responsible for recognizing concepts in "raw data", to categorization relying on perceived similarity, up to "higher" cognitive processes including communication through metaphors or creativity. It is definitively not to be ignored. But at the same time it is very difficult to study.

First of all, thanks to its multifarious aspects, it tends to be a slippery and hard to delimit notion. Many works (Indurkhya, 89) have made proposals to distinguish several types of analogies, emphasizing differences in purposes, a priori information and underlying processes. If some clarification results, at the price of complication, it remains to define precisely in each case both the goal of analogy (and the attached performance criteria) and the mechanisms involved.

Second, analogical reasoning is an unjustifiable (i.e. not logically valid) inference procedure. It goes beyond the deductive closure of the initial information and therefore cannot offer any warranty on its conclusions. But then what supports analogies ? What makes an analogy better than another one ? More concretely, why is it that it is so much used, apparently to the benefit of reasoning agents (as sanctioned by Evolution) ? Again, we encounter the problem of the evaluation criteria. More basically, the difficulty lies in the lack of firm referential system upon which to build and evaluate theories and models of analogy.

Responses to these problems have been twofold. One has been to seek some normative characterization of analogical reasoning whereby necessary conditions for sound inferencing are stated (Russell, 1987). Unfortunately this interesting approach so far has delivered very restrictive conditions that in effect exclude much of the subject matter. The other approach takes natural reasoning agents, prominently human ones, as standards against which to measure the quality of analogies and of the mechanisms that produce them. But of course, these natural yardsticks are subject to many parameters (perceived context, implicit goals, cultural background and so forth) that are impossible to securely control. Therefore this opens the door for endless arguments about the relevance and validity of each new experiment, and consequently of the tested models.

It is noteworthy that in this context, what is evaluated are not so much the end results of analogical inferencing, but rather the processes that are assumed to play a key role in their production. For instance, once it has been hypothesized that similarity judgments are at the

---

Keywords : Analogy, Theory and model for analogy, Machine Learning, Inductive learning theory, COPYCAT.

core of analogical reasoning (and many other cognitive processes as well), theories, models, and arguments center on similarity measurements and what they involve, in effect evacuating the fundamental question of why a high degree of similarity between a source case and a target case should entail highly reliable transfers of information from one to the other (leaving aside both the important issue regarding the objective nature of similarity (Medin et al.,1993) and the question of the modus operandi of these transfers).

This overall situation : a subject matter concerned with an inferencing process both presenting seemingly many different facets and manifestations, and inherently lacking sound justification, is reminiscent of the situation faced by the students of induction ten to fifteen years back. There also, there were plenty of models for inductive reasoning that were assessed on the face of their measured performance on chosen benchmarks, and a corresponding need for an established theory. The situation has changed recently (mostly thanks to Vapnik (1995), Valiant and many brilliant co-workers of the COLT (Computational Learning Theory) community).

This apparent aside on induction points out a third potential way of approaching analogical reasoning. Since it is supposed, rightly, that it is a core component of many cognitive processes, it should not be an isolated point with regards to its internal working and its performance criteria. In other words, properties and principles uncovered in studying other fundamental cognitive processes should hardly be expected not to be shared, at least in part, with analogical reasoning. Consequently, any theory and model of analogy should be consistent with theories and models for other, related, faculties. This could, and should, provide for good anchor points on which to erect models of analogy.

This is indeed the track that we take in this paper. In a way, we are pursuing a very ambitious goal, that of uncovering some fundamental traits that would constitute the basis for an overall theory that would encompass several cognitive faculties, including of course analogy making. We propose not to find justifications for analogical inferencing, an hopeless pursuit, nor to assess the value of one's model by comparison with natural reasoning agents, something necessary but not sufficient and never to be completely satisfactory nor convincing, but to present a theory of analogical reasoning that both satisfies a reasonable criterion for analogy, and at the same time is consistent with existing theories of inductive learning, a process that we argue is intimately related to analogical inferencing.

This paper presents the current state of this endeavor. Section 2 argues that analogical reasoning and induction are intimately connected while at the same time being different in important aspects. It also sums up the current state of accepted theories of induction. In section 3, we present our own model of analogy, showing in which respects it is intuitively appealing and how it maintains closed links with theories of inductive learning. Section 4 demonstrates on a canonical example that the model yields realistic results. Finally, section 5 sums up the state of this project and points to directions for future research.

2. ANALOGY AND INDUCTION : RESEMBLANCE'S AND DISSIMILARITIES

Deeply rooted in analogy surely rests the notion of similarity. At the least, analogy induces similarity, sometimes totally unexpectedly, as in creative analogy. The objective nature of similarity is the object of active debate within psychological circles (Medin et al. 1993), but it undoubtedly underlies categorization too : similar things tend to be grouped together in cognition. Analogical reasoning also shares many common points with induction, as we see now.

2.1A view on inductive learning and its theory

Figure 1 provides a flavor of what we are up to in inductive learning. A collection of examples, the learning set, is given, consisting of

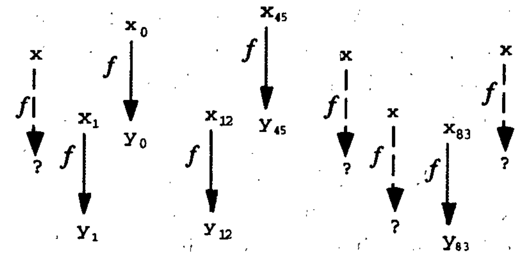


Figure 1. Inductive learning (in the supervised setting), consists in identifying a function f that "explains" the learning data (set of pairs $(x_i,f(x_i))$ and making the inference that the same f applies in unseen instances.

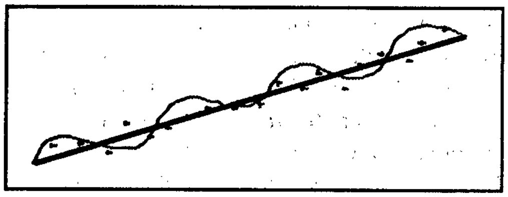


Figure 2. The best model for the data points is deemed to be the one that is at the same time "simple" and fits well to the data. Here, the linear model is simpler to specify than the polynomial one, and seems to fit equally well (or bad ?) the data points. Hence, following the MDLp, it should be preferred.

pairs $(x_i,f(x_i))$, and the goal is to infer what value would take the hypothetical and unknown function f on new points x_j. Generally, there is a cost associated with errors on $f(x_j)$, also called the risk, so that inductive learning consists in finding an hypothesis h such that the risk averaged[1] over the space of all possible instances, or the expected risk, be minimal.

Before the large diffusion of theoretical studies of induction (Vapnik,1995), the common view was that the obvious learning strategy was to select an hypothesis minimizing the risk over the learning set, called the empirical risk since it is measurable, in order to automatically get the optimal hypothesis with respect to the expected risk (one that by nature is unknown). This belief has been formalized and given a name : the Empirical Risk Minimization principle (ERM for short). In essence, what this principle states is that the best account for the learning instances is ipso facto the best one also for yet to be observed events. Vapnik, and many other theorists in the last fifteen years, have disproved this naïve view.

Of course, the philosophers knew this all along. There cannot be any miraculous basis for inducing general laws from specific observations. But theorists of inductive learning have gone further, specifying sufficient conditions for induction to be a reliable source of inferences. Sketched in broad lines, the now "classical" theory of induction states that induction is possible and reliable in proportion that the set of potential candidate hypotheses considered by the learner is restricted[2]. In other words, a learner that is able to explain any data set is hence unable to make induction, while a learner that can only consider severely restricted classes of concepts, if with these it may explain the observed data (available in sufficient quantity), is justified to generalize to other, as yet unknown, cases. Given that there is no "free lunch", the problem is now to chose a priori the right set of hypotheses.

It is noteworthy that, according to these theories, the confidence that one may put in inductive learning only depends on statistical quantities characterizing the hypothesis set taken as a whole, as well as the distribution and the number of learning instances.

Other theoretical approaches to inductive learning share this property. These are the bayesian perspective on learning and the related Minimum Description Length principle (MDLp). Roughly, they prescribe to select the hypothesis which is maximally probable given the observed data and their a priori probabilities (something that is easily computed with Bayes formula). The MDL view replaces this principle by one where one should chose the hypothesis such that the sum of its code length (within some well chosen coding schema) and the length of the description of the data encoded with the hypothesis be minimal (figure 2 illustrates this). It is a remarkable fact that it can be proved that the Vapnik theory and the MDL principle, starting from widely different premises, are nonetheless tightly linked. A fact that reinforces the confidence in these theories.

---

[1] More precisely, the averaging is weighted by the distribution over the instance space, so that more weight is given to dense areas, where it is more likely to encounter future events.

[2] Technically, these restrictions concern the possible partitions of the instance space that are induced by the hypothesis set. They are measured via statistical quantities, the most famous one being the Vapnik-Chervonenkis dimension.

This is all good and well, but does it have something to do with analogy ?

2.2 The same, yet different

As already noted, there are several types of analogies. Some involve the comparison of two given items (e.g. "abc" and "122333"), and some the completion of one item given the other (e.g. if "abc → abd", what should be the completion of "aababc → ?"). This last case (due to Hofstadter and his co-workers (Mitchell, 1993), (Hofstadter, 1995)) is a tricky one. We do not mean here that it might be difficult for the reader to infer the completion "aababcd", but that this is just a good example where one is made aware of the fact that much more has to be inferred. Indeed, nothing is given about the ways the strings (are they really ?) should be perceived, nor about the dependence relationship between "abc" and "abd" in the source case. Worse yet, the perception and interpretation of the source depends on the target probe. Had the last one be here "American Broadcasting Corporation → ?", that the source "abc → abd" would have been thought of completely differently. It is therefore evident that this type of analogy encompasses the former one where no completion, other than completion of interpretation, takes place. This is why we will consider this one type here.

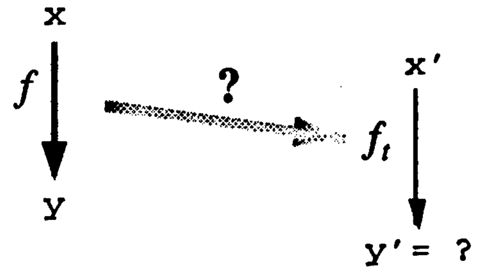


Figure 3. One view of analogy making enhances its inherent inferential aspect from limited information. Only x, f(x), and x' are known to the agent. From these "raw data", the agent must infer their interpretation, the dependence relation f in the source, and the corresponding "transported" dependence relation f_t in the target. From this follows y'.

If now, we take a look at figure 3, it may strike us that analogy is but a limit case of induction where one has access only to one learning instance. Under this perspective, analogy and induction are the same. And this is why we argue that surely their respective theories should be consistent so that they merge in between where very few learning instances are available.

On the other hand, there exist significant differences that make problematic the simple extension of the classical theories of induction to analogy, but also, as we will see, offer the perspective of refining these existing theories beyond their current state. Here is a list of these differences.

- The prediction is to be performed on one point only, not on the whole potential instance space. The notion of expected risk is therefore undermined to say the least.
- Each item potentially has its own referential frame (as in "abc → abd"; "122333 → ?", or better yet in "abc → abd"; "American Broadcasting Corporation → ?"). This is in contrast to the unstated assumption in induction that the looked for hypothesis **f** is the same all over the instance space.
- The target plays an important role in analogy, shaping the interpretation of the source, while it does not intervene in any ways in existing theories of induction.
- Finally, may be as a consequence of the above points, it is strongly believed that the "distance" between the source and the target plays a key role in analogy. In contrast, there is no notion of distance between instances in inductive learning[3].

To sum up at this point. We believe that the study of analogy should deliberately take into account related cognitive processes, such as categorization and induction, and try to make contact with the theories therein. This would more firmly anchor tentative theories and models for analogy. At the same time, developing theories adapted to the specific demands of analogy offers the perspective to refine the theories of the related cognitive process. To be more specific, incorporating the notion of distance between instances, and/or of local referential frames, into the theory of inductive learning, in needed harmony with theories of analogy, should result in finer theories of induction. Theories that, for instance, would better predict which amount of information is needed in order to be able to learn, say, some classes of concepts.

This is in accordance with this philosophical outlook that we have undertaken to look for a theory of analogy, one that would be faithful to the phenomena, and be related to theories of inductive learning.

3. A PROPOSAL

Let not be misled here, we are not, at this point, looking for the specification of a reasoning mechanism that would be a candidate for modeling analogy making, but we aspire to find a criterion for evaluating candidate analogies, a criterion that the best analogy should optimize. Recalling figure 3, it is clear that this criterion must depend on what is known to the reasoning agent, i.e. the source : x and f(x) (in the best of case including f itself), and the incomplete target : x'. It should also depend on prior knowledge which is the basis for the interpretation of the situations.

In addition to this, and following our policy, we should find a criterion that is consistent with the theory of induction. In particular, this criterion should take into account the "entropy" of the candidate hypotheses space, or, more intuitively, of the complexity of the candidate hypotheses. The idea being that the more underlying regularities are discovered in the data, the more its expression can be compressed. The MDLp is one expression of this general doctrine. We should therefore look for a measure of parsimony. The best analogy should correspond to the discovery of regularities both in the source and the target, regularities that should be as interrelated as possible. This last point being in agreement with a third desiderata : that the evaluation criterion reflects in some way our anticipation that analogy is linked to a notion of perceived similarity or distance between the analogs.

An evaluation criterion for analogy

In figure 4, we show how a version of the MDLp could be adapted to analogy. The best analogy should be the one that minimize the cost of the models or interpretations on which are based the perception of (x, f(x)) on the one hand, and, on the other hand, of x', while at the same time minimizing the cost of translating the interpretation of the source to the interpretation of the target. This is what is expressed in the following proposition.

Given M_S, M_T and f, it is easy to derive f_t by $f_t = pgm_{MS\text{->}MT}(f)$, that is the transformation of the expression of f within the referential associated with M_S by the program that transforms referential M_S to referential M_T. Then $f_t(x')$ may be computed.

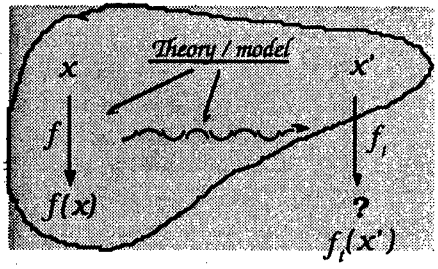


Figure 4.

---

[3] Beware not to confuse the notion of distance between instances, as in analogy, and between hypotheses or an instance and an hypothesis, as can be the case in induction.

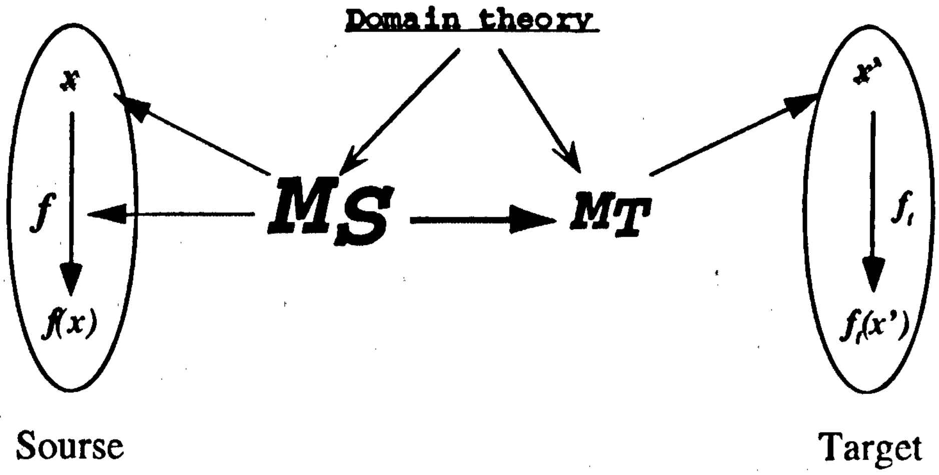


Figure 5. Following the theory presented here, any analogy involves interpretations or models, constructed from prior knowledge (the domain theory) that are local to the source : M_S, and to the target : M_T. From these, the specifics of each case can easily be reconstructed. At the same time, analogy making implies that a relationship be identified between M_S and M_T such that the two seem similar to each other. We submit that the best analogy is the one that minimizes the overall cost of specifying the models, there relationship (from M_S to M_T) and the derivation of the specifics of each case.

Proposition :

The set of models and descriptions M_S, M_T, x, f, x' that minimizes the formula[4] :

Total_length = $L(M_S) + L(x|M_S) + L(f|M_S) + L(M_T|M_S) + L(x'|M_T)$ is the one associated with the best analogy between the source and the target.

4. ILLUSTRATION

This section intends to illustrate the above conceptualization. It is not meant to demonstrate its value as a model of the human ability in making analogies. This is beyond the scope of this short paper, and would require a careful discussion of representation primitives, suitable coding system, and hypothesized prior knowledge.

4.1 The domain

In order to keep things manageable, we have chosen a domain where it is easy to define representation primitives and theories, and yet which presents enough richness to be demonstrative of the wealth of issues in analogy-making. This domain is inspired from the microworld developed by Hofstadter et al. for the COPYCAT project (Mitchell, 1993).

The basic objects in this world are the 26 letters of the alphabet, but it would be straightforward to add numbers or geometrical shapes. The task consists in finding how a letter string is transformed given, as an example, another string and its transform. For instance, given that abc => abd (the source), what becomes of **iijjkk** => ? (the target). The problem, quite familiar in IQ like tests, is thus to identify the relevant aspects and transformation at work in the source that can best be mapped to the target problem. It is very easy to make up a whole variety of problems that test the range of analogy-making.

---

[4] L is taken to be a function measuring the cost or length of its argument expressed in bits. We do not dwelve here in technical details about what that involves. We refer the reader to (Li & Vitanyi,1993) for a thorough introduction to algorithmic complexity theory on which our model is based.

- Features describing the conceptual structures :
  - orientation (-> / <-) — *1 bit*
  - cardinality or number of elements : n — *$\log_2(n) + 1$ bits*
  - length : l — *$\log_2(l) + 1$ bits*
  - starting or ending with element = x — *L(x) bits*
- **Letter** — *(1/2)*
  Particular letter (e.g. 'd') — *(1/2.26)*
- **String** (orientation,elements) — *(1/8)*
  L = 3 + L(orientation) + _ L(elements)
  e.g. L('a3bd' with orientation = ->) = 3 + 1 + $\log_2((1/2.26)^3$ + L(3)
  = 3 + 1 + 18 + 3 = 25 bits
- **Sequence** (orientation, type of elements, succession law, length, starting or ending with) — *(1/8)*
  L = 3+ L(orient.) + L(type) + L(law)+ L(length) + L(start/end)
- Description and length of a **succession-law**
  succ(type-of-el.,n,x) _ the nth successor of the elt. x of type-of-el.
  L = L(type) + L(n (see below)) + L(x)
  L(n) = L(1/6) if n=1 or -1 (first successor or predecessor)
  L(1/3) if n=0 (same element)
  $L((1/3).(1/2)^p)$ otherwise (with p=n if n³0, p=-n otherwise)
- First / last — *1 bit*
- nth — *n bits*

Table 1. List of some representation primitives with their associated description length either in bits or defined as probabilities.

Hence, the string **abc** could be described as:

'abc' _ **String** (1/8)
orientation : -> (1/2)
1st='A', 2nd='B', 3rd='C' $(1/4.26)^3$
TOTAL Length : **21 bits**

or else as :

'abc' _ **Sequence** (1/8)
orientation : -> (1/2)
type of elements = letters (1/2)
succession-law :
succ(elt(letter=x) = elt(succ(letter,1,x))
L(letter) + L(1st succ) + L(x)
= L(1/2.1/6.1) = 4 bits
length = 3 3 bits
starting with element(letter='A') (1/26)
TOTAL Length : **17** bits

Following (Mitchell, 1993), the background knowledge or domain theory includes the basic representation primitives and the conceptual structures that allow to describe and highlight various aspects of the situations at hand (see table 1). In order for the quality criterion to be computable, each construct is associated with a number, that corresponds either to a prior probability from which it is easy to draw the related length using the relation L=-$\log_2$(P) (e.g. the concept of string is associated with the prior 1/8, hence is of length 3 bits), or directly to a length in bits (e.g. the concept of nth requires n bits). These numbers can be modified either manually or through learning to yield various biases corresponding to a variety of contexts or prior knowledge.

It is clear that the last description, which more fully represents the structure of the string abc, is the most economical one, even though it describes it more completely than the first description which corresponds to the perception of a set of three letters.

4.2 Experiments

We have tested the above scheme on a variety of analogy problems in order to see what rankings the criteria would give to various possible solutions. Limited space prevents us from giving a full account of the derivation of the complexity figures. The overall method is as follows. For each pair (Problem; Solution), we hypothesize associated models or perceptions. For instance, iijjkk can be perceived as a string of letters, or alternatively as a sequence of successive pairs of letters. Then, a program computes the algorithmic complexity of these constructs and of the transformation programs that allow to derive one description from another. The associated figures are reported in table 2.

Problem: abc => abd ; **iijjkk** => ?

Solutions :

S1 : "Replace rightmost group of letters by its successor" **iijjkk => iijjll**

S2 : "Replace rightmost letter by its successor" **iijjkk => iijjkl**

S3 : "Replace rightmost letter by D" **iijjkk => iijjkd**

S4 : "Replace third letter by its successor" **iijjkk => iikjkk**

S5 : "Replace Cs by Ds" **iijjkk => iijjkk**

S6 : "Replace rightmost group of letters by D" **iijjkk => iijjd**

	S1	S2	S3	S4	S5	S6
$L(M_S)$	10	9	11	11	12	11
$L(x\vert M_S)$	8	18	18	18	22	15
$L(f\vert M_S)$	4	4	3	7	8	3
$L(M_T\vert M_S)$	5	0	0	0	0	17
$L(x'\vert M_T)$	8	36	36	36	42	15
Length (bits)	**35**	**67**	**68**	72	**85**	**62**
Rank	1	3	4	4	6	2

***Table 2** The figures corresponding to the evaluation formula are reported for various solutions to the problem considered. Solution 1 emerges as a clear winner, which is also the choice of most human subjects when asked to rank these solutions.*

5. CONCLUSION AND PERSPECTIVES

These experiments and calculations, cannot and do not pretend to be conclusive. They rely on many hunches and simplifications that would need to be more carefully set. Indeed, it is natural that such be the case, since this proves by the same token that our model nicely incorporate contextual effects and the possibility of learning (concepts and associations), and of the consequences these may have on analogy making. Still, these results show that the proposed scheme does not seem entirely unreasonable from the point of view of a comparison with natural cognition. But we also believe that most promising is the fact that this model is tightly linked with induction theory. Nonetheless, it remains unclear why a high degree of similarity, or the possibility of a simple interpretation of the analogs lends credit to the analogical inference. This is a question we actively study.

Else, one of our current research project is to better ground our calculations on the theory of algorithmic complexity, to maintain close links with inductive theory, while at the same time experimenting with many more examples from a variety of domains. We also study how mechanisms for the actual production of analogies (not only for evaluation) could be derived from this perspective.

6. REFERENCES

Hofstadter D. (1995). Fluid Concepts and Creative Analogies. Basic Books.

Indurkhya B. (1989). Modes of Analogy. LNAI-397, Springer-Verlag.

Li & Vitanyi (1993). An introduction to Kolmogorov complexity and its applications. Springer-Verlag.

Medin, Goldstone & Gentner (1993). Respects for Similarity. Psychological Review, 1993, Vol.100, No.2, 254-278.

Mitchell M. (1993). Analogy-Making as Perception. MIT Press.

Russell S. (1989). The Use of Knowledge in Analogy and Induction. Pitman Publishing and Morgan Kaufmann, 1989.

Vapnik (1995). The Nature of Statistical Learning Theory. Springer-Verlag.

JUSTIFICATION OF ANALOGY BY ABSTRACTION

Hiroaki Suzuki
Department of Education,
Aoyama Gakuin University
Shibuya 4-4-25, Tokyo
150-8366, Japan
susan@ri.aoyama.ac.jp

INTRODUCTION

How is reasoning by analogy justified? Why can we map non-identical elements in the source and target analogs? Is it valid to transfer some elements in one analog to another? Although the problem of justification is of critical importance for the research on analogy, only a few studies have discussed them seriously (Gentner, 1983; Indurkhya, 1992). The aim of this paper is the developing of a new framework that provides an answer to the justification problem.

In the real world, analogical reasoning is widely used and has strong power in various kinds of human activities, such as problem-solving, learning, discovzotherapy, literature, myth, political and legal argument (Holyoak & Thagard, 1995). It is a powerful tool for providing a solution, creating a new idea, arguing against, and persuading opponents, making ideas more explicit and impressive.

However, some blame analogy and claim not to use it, because analogy is known to be a dangerous mode of reasoning. Analogical reasoning, like induction, does not have logical validity. Actually, there are abundant examples of misuse of analogies in various kinds of human activities, such as education, science, political arguments, commercial advertisement (Gentner & Jeziorski, 1993; Holyoak & Thagard, 1995; Indurkhya, 1992). More depressing findings were obtained by Chi and her colleagues (Chi et al., 1989). Their study found poor learners' excessive reliance on analogy. These learners frequently looked back to previous problems, read them extensively, tried to map them to the current problem, which resulted in poor performance on transfer tasks.

What was mentioned above shows two opposing pictures. Analogy enriches human cognition and gives new insights in some cases. In other cases, analogy obscures our rationality and falls into poor learnersÕ desperate heuristics.

The purpose of the paper is to develop a new framework to give explanations of how reasoning by analogy is justified, and in what condition analogies are, at least, psychologically valid.

In the next section, I analyze the conditions for justified analogies. According to the analysis, analogy should be treated as a kind of categorization. This means that analogy is a ternary relation between the base, target, and their superordinate category (abstraction), rather than a binary relation between the base and target. Second, I will show that this formulation greatly reduces the computational complexities in retrieval and mapping. Third, I will try to figure out the characteristics of categories by the findings obtained from an informal observation. Finally, I will reexamine the relationships of analogy to other kinds of cognitive activities, based on the proposed framework.

JUSTIFICATION

Identic ality

Although controversies are still continuing about many aspects of analogy, there is one basic assumption that few deny. This assumption is that analogy involves mapping from the base to the target. A set of elements in the base is corresponded to a set of elements in the target. Another set of elements in the base are, then, transferred to the target to create new inferences.

Here, one can ask why some elements in the base can be mapped and transferred to the target. What enables mapping and transfer between the base and target? It is a difficult problem, but Leibnitz gave a partial answer to this problem. According to his principle of the identity of the indiscernibles, if two things are identical, any of their predicates can be transferred. This principle suggests that mapping requires identicality between the base and target.

However, as long as analogy is concerned, this principle is too rigorous to be applied , because a base is not identical to a target, by definition. The base is represented qualitatively different from the target. They are in no way identical.

Categorization

Thus, it is necessary to find a cognitive mechanism that makes two different things identical in some respects. Logically, it is impossible, but there is one psychological mechanism that can do it approximately. It is categorization. If two things belong to the same category, they are properly said to be identical in terms of the category. Suppose, for example, that there are two cats that differ in their size, color, etc. Despite of these differences, they are identical with respect to their Òcatness.Ó If two cats belong to the same category, attributes and predicates important with respect to the category can be mapped from one cat to the other.

The same argument can be applied to the theory of analogy. If there is a superordinate category whose members are the base and target, they are properly said to be identical, with respect to the category. Thus, properties and relations shared by the base and the category can be transferred to the target.

The discussion so far leads us to change the basic framework of analogy. As I said earlier, analogy has been considered to be a binary relation between the base and the target. However, if the argument above is correct, it follows that we should consider analogy as a ternary relation between the base, target and their category. But, the term "category" usually refers to a preexisting taxonomic category, such as animal, plant, dog etc., so I introduce a more neutral term, *abstraction*, here. Note that the term abstraction here refers to an abstracted mental entity, not to the action to abstract.

Attempts have been made to incorporate abstractions to the theory of analogy. In artificial intelligence research, several models of analogy have made explicit use of abstraction (Greiner, 1988; Kedar-Cabelli, 1985; Russell, 1988). Glucksberg and Keysar (1990) and Lakoff (1993) assume abstracted mental entities in understanding metaphorical statements, although there are controversies between them. A number of studies on transfer of learning have shown the importance of abstractions (see, for example, Gick & Holyoak, 1983; Goswami & Brown, 1989). Thus, the framework proposed here is not a new one. Rather, my attempt should be considered as a synthesizing one.

COMPUTATIONAL CONSTRAINTS

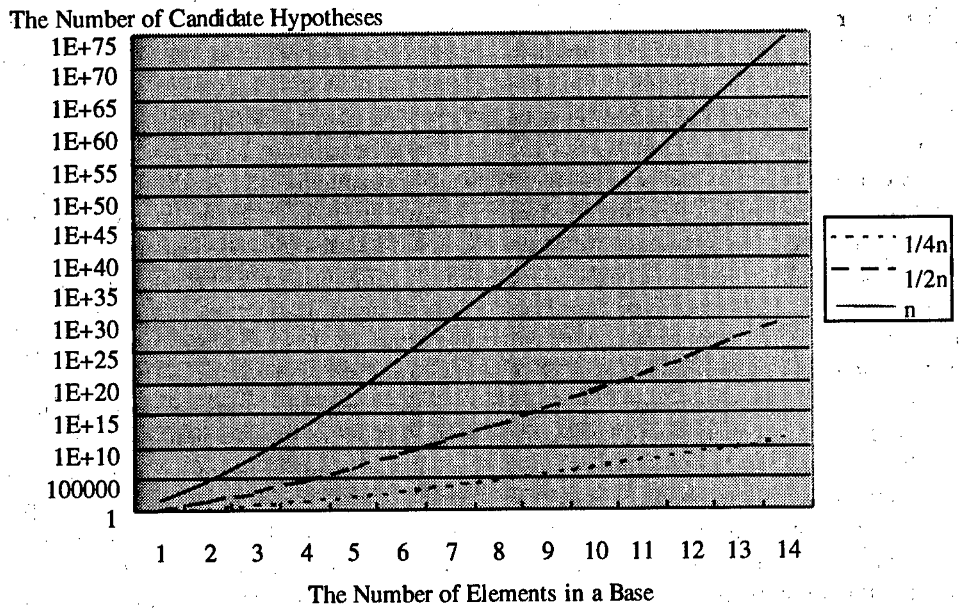
By introducing the notion of abstraction, we obtain a couple of constraints that greatly reduce the computational complexities in retrieval and mapping.

Retrieval

An important consequence of introducing the abstraction is that the analog retrieval mechanism can make use of hierarchy. Not a few researchers admit that our long-term memory is represented hierarchically from most concrete to most abstract ones.

If the retrieval mechanism makes use of the information about the hierarchy, the cost of retrieval is obviously reduced. For example, if an abstraction is judged to be irrelevant in the process of categorization, an analogizer needs not consider all of its descendents. Theories ignoring the hierarchical information have to test every subcategory even after its ancestral abstraction is rejected.

There is another benefit. The more ascending a hierarchical tree, the less information is



available. Consequently, one may sometimes descend the tree to obtain further information. In this case, the hierarchical structure constrains further search. If you select an abstraction at some level and want to get more information, you need not search the entire space. Instead, it is sufficient to search items that are descendents of the abstraction previously selected.

One of the problems to be considered here is whether concrete base analogs are hierarchically organized. Many agree with the hierarchical organization of the common natural kinds, but how about knowledge structures used in analogy?

Memory organization is found in more complex materials such as stories and episodes. Although there are controversies, some researchers showed that the story grammar type of knowledge structure constrains encoding and retrieval of stories.

The second line of evidence comes from Shank's Mops and TOPs type of knowledge organization. Reflecting Black, Bower, & Turner's experiment, Schank elaborated his theory of scripts to include more abstract knowledge structures. According to him, there are knowledge structures that hierarchically organize concrete representations of specific events. They are called, MOPs, metaMOPs, universal MOPS. In addition, he assumed a different kind of structures that organize thematically similar events, and he called it TOPs (thematic organization packets). He believed that these best explain cross-contextual reminding.

The third line of evidence comes from Fukuda's work. In his experiments, subjects' reminding was greatly improved when they were given cues at the moderately abstract level, compared with when given very similar stories as cues. The superiority of such a cue strongly supports the idea that there exist abstractions and that concrete episodes are organized around the abstraction.

Mapping

In the mapping process, abstractions make two contributions, both of which reduce the computational costs involved in mapping. The first one is concerned with the selection of candidate elements to be mapped. Suppose that a base has *n* elements. The number of the candidate sets to be mapped amounts to $2^n - 1$. This obviously causes combinatorial explosion if *n* is getting larger.

However, if an abstraction is involved in mapping, one need not suffer from it. It is because what is true for the abstraction must be true for its subordinate target. It follows that every element in the abstraction can be, and should be, mapped. Although one still has to decide which element in the base correspond to which element in the target, the computational gain in the selection of a candidate set is very large.

The second benefit is concerned with the number of elements in the abstraction. The number of elements in an abstraction is, by definition, smaller than that of its subordinate, concrete base analogs. It is impossible to make a general estimation of how much smaller the elements in the abstraction is, but the reduction in the number of elements produces huge computational gain in many cases.

For example, if one maps *n* elements of the base to the target, the resulting number of possible mappings is the permutation of *n*, shown by the thick line in the graph. As you see, it is approximated by an exponential function. Suppose that an abstraction has a half of the elements. The number of candidate hypotheses is depicted by the broken line (the dotted line shows the number of hypotheses when the number of the base element is reduced to a quarter of *n*). Although the number of possible mapping hypotheses is exponential even assuming the abstraction, the computational gain is huge compared with the cases without abstractions.

ABSTRACTIONS IN ANALOGICAL REASONING ABOUT ELECTRICAL CIRCUIT

Informal observation

This is the stage for the present framework to be more concrete. My favorite example is people's natural reasoning about the electric circuit. As Gentner & Gentner (1983) reported what type of base analog is used affects subjects prediction about the behaviors of the circuit. When a water flow system was introduced, subjects correctly infer the change of the electricity when a battery is added serially. On the other hand, subjects prediction improved in the case of parallel resistance, when a teaming crowd analogy was taught.

In the experiment, they gave subjects either analog explicitly, and asked them to use it when answering the problems. However, people can draw analogies spontaneously even without such instruction. From my observation, most university students used liquid flow analogies initially, although they were not exactly the same as the water flow, as I will show you later.

Such naturally drawn analogies tell us many things. First of all, although most subjects used a kind of liquid-flow analogy, it is very dubious that their analogies were based on a specific experience about a water flow system. It is hard to imagine that they had seen water flowing in the closed circuit with a pump, even harder to imagine they had seen a parallel circuit with two pumps attached serially! If they did not have any experience with it, how could they make analogies? This shows that in naturally drawn analogies, the possibility of making use of very concrete, episode type of base analog is very low.

Second, the mapping was very immediate, so immediate that they seemed not to be in trouble with candidate mapping hypotheses. From my observation, no single case was found that they made mistake in finding correspondence. Essential parts in the base and target were immediately mapped, while non-essential parts seemed not to be even for a slightest consideration. The protocol shows no statements such as pump's having a lever or a switch, pumps needs of external forces, although they play causal roles in the actual water flow system.

Third, we observed the on-line construction of a base. When subjects were asked to estimate heating values at resistance, many subjects spontaneously and naturally switched the source analog from the liquid-flow to the particle-flow. That is, they changed the flowing entity from liquid to something solid, such as people, small stones, or particles. The shift seems to be done because water was judged not to be a relevant analog for the generation of heat.

These solid objects, instead, enable people to naturally infer the generation of heat by the friction of contacting parts.

Flowing system abstraction

The picture drawn by the observation is quite different from the ones that the current models of analogy do. Despite of the unavailability of concrete base analogs, people had little difficulties in reasoning analogically about the behaviors of the electric circuit. This fact suggests that an abstraction, a flowing system, is responsible for subjects' analogical reasoning. This abstraction is very simple in the sense that it consists of only three components: a flowing entity, path, and force. A typical relation between them is that the force causes the entity to flow through the path.

The simplicity of the abstraction partly explains the immediate mapping. Since there are only three components that are distinct, and every component of the abstraction is applied by definition, there are little possibilities for misunderstanding the mapping relations.

The flowing system abstraction is a higher-order abstraction, in the sense that every component is variable. Thus, it must be supplemented and enriched by contextual information involved in the problem situation, when it is actually used. This enables abstractions to be flexible. Even when people cannot access to a concrete base analog, they can naturally make useful inferences, by instantiating the abstractions under the constraints posed by the problem situation.

Furthermore, these characteristics explain the on-line construction of a new base analog. As I reported earlier, subjects could easily shift from the liquid-flow to the particle flow analog by changing the flowing entity when they dealt with the generation of heat. The ease of the shift cannot be explained, without assuming the flowing system abstraction. If people had used an actual water flow system as a base analog, the shift should not have been done so easily. This is because people have to retrieve a new analog by examining all the candidate analogs again, and they have to replace every component of the analog with new ones: a water pump with a loud speaker, a pipe with a road, etc. However, if one assumes the abstraction, the search for a new analog is constrained. Furthermore, it is enough to change one of the components of the abstraction, because the existence of the pushing force and the path is guaranteed by the abstraction.

In addition, the flowing system abstraction provides the global coherence when changing the analogs. If one uses a completely new analog, there is a possibility of inconsistency between what have been inferred and what will be inferred. On the other hand, inferences based on old and new analogs are consistent if they are descendents of the same abstraction. In the case of the electric circuit analogy, inferences based on the liquid flow analog are guaranteed to be consistent with those based on the particle flow analog.

Some researchers have emphasized the process of adaptation in analogy. Since there are few problems that a base analog can directly be applied, it is often necessary to adapt analogs to the current problem situation. For a flexible adaptation, it would be better source analogs to be small and simple, like the flowing system abstraction. It is difficult to modify and adapt big, deep, and complex analogs that contains a lot of information.

Contrasting abstraction-based view with current theories of analogy

The framework proposed here contrasts sharply with that of the dominant theories of analogy. According to the dominant view, episodes are represented almost literally in the form of first-order predicate logic. Since no abstraction or summarization is assumed to take place in encoding source episodes, each source episode forms a large, deep, complex structure. In addition, each analog is stored in a relatively isolated fashion. Thus, some assume only surface level matches (Forbus et al., 1995), while others can only make use of word-to-word level relations (Thagard et al., 1990). In mapping, many theories share the assumption that initial mapping is carried out syntactically. Since this type of mapping generates a large number of mapping hypotheses, one or more constraints are called for to reduce them (Falkenhainer et al., 1989; Holyoak & Thagard, 1989). The mapped structure is static and isolated in the sense that it is prestored in the source analog and has few relations to other analogs. Thus, when shifting a source, an analogizer has to reiterate the entire processes.

On the other hand, the abstraction-based view of analogy assumes small, simple abstracted mental entities as source analogs. A small number of variabilized components are involved in abstractions. Each source abstraction is connected to form a hierarchy. In mapping, variable bindings or unification take place, in a deductive fashion. Since an abstraction involves a small number of distinct elements, the number of possible mapping hypotheses is small. The resulting structure is liable to modification under the constraints posed by the target analog and task goal. In this sense, analogy by abstraction is dynamic and constructive.

The findings obtained from the informal observation of people's spontaneous analogical reasoning are not compatible with the dominant view. First, there seem to be no large, complex source analogs available. Second, people retrieved the source analog very rapidly. It seemed that only a limited number of candidate analogs were in consideration. This suggests that subjects may make use of the hierarchical information in retrieval. Third, mapping was rapidly carried out without mistakes. This suggests that they did not suffer from a large number of mapping hypotheses, which in turn leads us to the idea that a source analog actually used did not have a large, complex structure. Finally, subjects shifted from one source to another flexibly and naturally, by changing a part of the source analog. It would have taken relatively long time if they had replaced the original analog with a completely new one. This indicates that they did not use an analog representing the actual water flow system.

These findings are best explained by the abstraction-based view of analogy, which assumes small, simple variabilized mental entities connected hierarchically.

RELATIONS TO OTHER KINDS OF COGNITION

A number of researchers have explored the processes and structures of analogy for many years. They have revealed what subprocesses are involved in analogical reasoning, what affects human analogy making, as well as how and where analogies are used. These findings lead to computational theories of analogy. By their competition, the levels of analysis have been greatly improved, which in turn leads to greater sophistication of the theories.

However, the relationships of analogy to other kinds of cognition have been missed in the course of the scientific endeavor. Analogy plays a central role in human cognition, but it seems strange that there is a cognitive engine designed specifically for making analogies. It might be that analogy is a special combination of more basic cognitive components. If so, we should explore the relationships of analogy to other kinds of cognition.

The proposed framework opens the door of analogy to other kinds of cognition. In this section, I briefly review the relationship of analogy to categorization and deduction.

Categorization

One important relation is to categorization. Although abstractions accessed in the course of analogical reasoning are different from common categories, the underlying mechanisms are the same. Both assume the hierarchical structure and the inheritance of properties.

Certainly, dominant models of categorization seem to be a little bit too simplistic, because they do not have principled methods distinguishing structural and surface information. Thus, as Ramscar and Pain (1996) pointed out, the model of categorization should be modified and enriched by the findings obtained from analogy research.

Deduction

A striking finding provided by the framework is that analogy is similar to deduction. My proposal is the following: given a target is a member of an abstraction, and that abstraction has a property X, then the target has the property X. This form of reasoning is properly said to be a categorical syllogism. We explain why some analogies seem to be psychologically valid. This is because they are deduction.

However, I do not intend to reduce analogy to deduction. My position is the opposite. From my viewpoint, deduction is a kind of analogy. Abstractions used in the processes of analogy do not have the same status as premises in deduction. People may induce a wrong abstraction in some cases, while they may access to a wrong abstraction in other cases. Thus, there exists uncertainty in analogical reasoning. On the other hand, categories appeared in deduction are fixed, and proved to be relevant. Thus, no ambiguity is found in deduction. If you admit the discussion above, you will notice that deduction is a special case of analogy, not vice versa.

ACKNOWLEDGMENT

Many thanks to Bipin Indurkhya and Mark Keane for their helpful comments on the talk based on the earlier version of the paper.

REFERENCES

Chi, M. T. H., Bassok, M., Lewis, M. W., Reiman, P. & Glaser, R. (1989) Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Psychology*, **13**, 145 - 182.

Falkenhainer, B., Forbus, K. D. & Gentner, D. (1989) Structure mapping engine: Algorithm and examples. *Artificial Intelligence*, **41**, 1 - 63.

Forbus, K. D., Gentner, D., & Law, K. (1995) MAC/FAC: A model of similarity-based retrieval. *Cognitive Psychology*, **19**, 144 - 205.

Gick, M. L. & Holyoak, K. J. (1983) Schema induction and analogical transfer. *Cognitive Psychology*, **14**, 1 - 38.

Gentner, D. (1983) Structure-mapping: Theoretical framework for analogy. *Cognitive Science*, **7**, 155 - 170.

Gentner, D. & Gentner, D. R. (1983) Flowing waters or teaming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.) *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. & Jeziorski, M. (1993) The shift from metaphor to analogy in Western science. In A. Ortony (Ed.) *Metaphor and Thought*. Cambridge, UK: Cambridge University Press.

Glucksberg, S. & Keysar, B. (1990) Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, **97**, 3-18.

Greiner, R. (1988) Learning by understanding analogy. *Artificial Intelligence*, **35**, 81-125.

Goswami, U. & Brown, A. L. (1989) Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition*, **35**, 69 - 95.

Holyoak, K. J. & Thagard, P. (1989) Analogical mapping by constraint satisfaction. *Cognitive Science*, **13**, 295 - 355.

Holyoak, K. J. & Thagard, P. (1995) *Mental Leaps: Analogy in Creative Thought*. Cambridge, MA: MIT Press.

Indurkhya, B. (1992) *Metaphor and Cognition: An Interactionist Approach*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Kedar-Cabelli, S. (1985) Purpose-directed analogy. In Proceedings of the Seventh Annual Conference of the Cognitive Science Society, 150 - 159.

Lakoff, G. (1993) The contemporary theory of metaphor. In A. Ortony (Ed.) *Metaphor and Thought*. Cambridge, UK: Cambridge University Press.

Ramscar, M. & Pain, H. (1996) Can a real distinction be made between cognitive theories of analogy and categorisation? In Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society. 346 - 351.

Russell, S. W. (1988) Analogy by similarity. In D. H. Helman (Ed.) *Analogical Rea-*

soning: Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy. Dordrecht, Netherlands: Kluwer Academic Publishers.

Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990) Analog retrieval by constraint satisfaction. *Artificial Intelligence*, **46**, 259 - 310.

ANALOGY MAKING AS A CATEGORIZATION AND AN ABSTRACTION PROCESS

Emmanuel Sander and Jean-François Richard
Department of Psychology, CNRS ESA 7021
University of Paris 8
93526, Saint-Denis, France
sander@univ-paris8.fr richard@univ-paris8.fr

ABSTRACT

An object can be categorized in a rather arbitrary number of ways. A shoe can be categorized as a Nike, model X, size 44 (by the person making an inventory of a sports shop), as a Nike (by the customer), as a sports shoe (by somebody who intends to go jogging), as a shoe (by somedy who looks for shoes), as something used to go from place to place, as a covering that comes in direct contact with the terrain, etc. A tire can be categorized as a snow tire, as a tire, as something used to go from place to place, as a covering that comes in direct contact with the terrain, etc. What happens when drawing an analogy between a tire and a shoe (example from Gentner, 1988)? We will argue that, in many cases, analogy can be viewed as a categorization process within a network of categories. It might be a straightforward categorization, in which case the first category selected (source) is relevant for drawing the analogy. In this case, it could be debated whether this process should be called analogy or categorization. However, many cases considered as analogy in the literature can be plausibly described as situations of categorization. It might also require an upward search in a network of categories. The difficulty of an analogy can then be due to the difficulty of this search, either due to the number of steps required to categorize at the appropriate abstract level or to the difficulty to access that level.

In this paper we will first describe the semantic network within which we assume that the analogy is drawn. We will then discuss how some analogies can be seen as categorizations and how other analogies can be seen as involving an abstraction process.

A SEMANTIC NETWORK OF CATEGORIES, MEDIUM OF THE ANALOGY

Before intoducing the semantic network, we present some results and conceptions about categorization that we will rely on.

Some results and conceptions about categorization

There is a tendency to categorize objects at their basic level (Rosch, 1978). However this categorization is not systematic. It is influenced by the context and in particular by the goal of the categorizer (for instance an apple can be seen as a fruit but also as a thing to take for a picnic, Barsalou, 1991), and influenced by her or his expertize in the field (experts in birds will not categorize them at the same level than lay people, Tanaka & Taylor, 1991).

The inclusion relation is a key relation within categories (Smith & Medin, 1981). Yet, a network of categories can not be seen as a tree (a taxonomy). As Richard and Tijus (1998) suggested, a tomato can be seen as a fruit or as a vegetable depending whether the context is preparing a dinner or a course of biology. The structure which organizes categories is more probably a kind of hierarchy in which a subordinate can have several superordinates, that is a lattice (Poitrenaud, 1995).

Categories do not only apply to group of objects and a word of the language is not needed to designate a category. There exist natural and artifactual categories (Rosch, 1978, e.g., birds and T.V.), categories of scenes and environments (Tversky & Hemenway, 1983, e.g., fancy restaurant), categories of events (Morris & Murphy, 1990, e.g., food shopping), goal oriented categories, (e.g.: 'things to take on a camping trip') which can be ad hoc categories, that is built for the need of a task but can also become well established in long term memory (Barsalou, 1991). For instance 'thing that is desired but can't be obtained and hence is denigrated' (see below) or 'situations in which an action taken to remedy a problem actually defeats the main purpose of the thing affected by the problem' (Mitchell, 1993) might be considered as categories.

Categorization is often viewed as a classification tool (Rosch, 1978). However, categorization is not only a cognitive economy device allowing people not to deal with too many different objects in the world. Categorization is used to infer non-evident properties from evident properties. If we consider that an object belongs to a category, we can predict more about this object that what we actually observed (Anderson, 1991). When we see a shoe, we can infer that it can be used for walking, that it will be damaged if we use it without taking care of it, that it might have been made by children in South-East Asia.

The properties of an object belonging to a category are not equally easily accessed. While some properties are context independent, others are activated only in specific contexts (Barsalou, 1982). For instance, a basketball rolls is activated in any case, but that a basketball floats is activated in specific cases like basketballs charged on a boat. This context dependency can be interpreted as an access to a superordinate category depending on the context (for instance, depending on the context, a basket ball might be seen as belonging to the category of floating objects, Sander, 1997).

The categories can have a complex structure. The views that categories are represented by a list of features (feature models) or by some instances (exemplar models) have been challenged. Several authors consider that our representation of concepts is structured in a more complex way and might include relations between features or with other concepts (Murphy & Medin, 1985; Wisniewski, 1995). For instance, in our representation of the concept car, we have probably information about the respective roles of the different parts of a car (the wheels, the seats, etc.) and about how these parts interact.

Description of the semantic network

Once we face a new situation (a target), we claim that analogy making can be described as a search and property attribution mechanism, which operates within a semantic network which has been activated by this target situation. The construction of this semantic network is circumstantial and contextualized: it is done within the context of a task, in the same way as the construction of the ad hoc categories (Barsalou, 1991). The semantic network includes semantic and functional knowledge associated not only with the objects present in the situation, but also with the objects and categories associated to the situation: the semantic network, medium of the analogy, is a part of a general knowledge network seen under a certain point of view, that of the task. It is built from two operations of selection: an operation depending on the nature of the objects of the situation and an operation depending on the task and on its constraints. As the selection is made also at the level of the goals, this implies that the same device, used for two different tasks, might not generate the same semantic network. In particular, this leads to a selection among the potential superordinates. A computer, used as a word processor will evoke the typewriter, the domain of writing, and the domain of manipulating objects (Sander & Richard, 1997). The same computer used to deal with a data base evokes (at least in France) the wellspread device known by the name of Minitel which includes a keyboard and a screen and is used to give access, via the telephone, to many services. Despite the fact that all this is done through the use of the keyboard, other domains are evoked (the domain of the telephone and the domain of communication, Richard & Tijus, 1998).

Within the formalism that we use (PROCOPE, see Poitrenaud, 1995), categories, considered from the point of view of the inclusion relation, are the nodes of a network; the links between the nodes represent the relation «is a kind of». From the point of view of the part-whole relation, different properties can be associated with each part (e.g., a wheel of a car has some properties and a seat has other properties). Properties are associated with a category, those which are activated when an object is considered as belonging to the category, as Anderson, 1991 considers. One specificity of this formalism (Poitrenaud, 1995) is that goals that can be achieved on an object are considered as properties of the object (for instance a property of a shoe is that it can be used for walking, a property of a word, considered as an object of a text editor, is that it can be moved).

Once such a semantic network has been activated, two kind of analogies may be distinguished. Those which can be seen as straightforward cases of categorization and those involving an abstraction process.

ANALOGIES THAT CAN BE SEEN AS STRAIGHTFORWARD CATEGORIZATIONS

Analogy as a categorization process

Several investigators have already claimed that there is a continuum between analogy and categorization (e.g. Hofstadter, 1995; Turner, 1988). Actually, in both analogy and categorization, a known situation (the category or the source) is used to treat a new situation (the object to be categorized or the target) as if it were familiar (Holyoak & Thagard, 1995; Spalding & Murphy, 1996). In both cases, one of the main purposes is inferential: the knowledge about the source or the category is used to infer features of the target (Holyoak & Thagard, 1995; Anderson, 1991).

Analogy is, in some cases, a straightforward case of categorization in which the target situation is assimilated to a reference class, which is the source. Properties common to the source and to the target are used to access the source, which enables properties belonging to the reference class (the source), to be attributed to the new situation. The basic process in this case is the search for a relevant source. We consider that a source is accessed according to the salient features it shares with the target (Vosniadou, 1989). The salient features are those which the participant accesses in the new situation, taking into account her or his knowledge and the context (Kokinov & Yoveva, 1996), and that she or he considers as relevant.

In our study of learning text editing (Sander & Richard, 1997), we have shown that, in a first step, the typewriter is a source of analogy for the participants. We found that all the procedures imported from the typewriter were used by all the participants in the experiment from the beginning of a learning session, whereas only 12% of the procedures which were not direct adaptations of typewriter procedures where used by them. The analogy can be described this way: the text editor is categorized as a typewriter, as this is the known domain which shares the greatest number of salient features. The general goal is the same: to type a text, and objects are shared: a keyboard and a surface on which what is typed appears (screen or sheet of paper). Knowledge associated to the actions that can be performed with a typewriter is described in a schematic way by the network of the Figure 1.

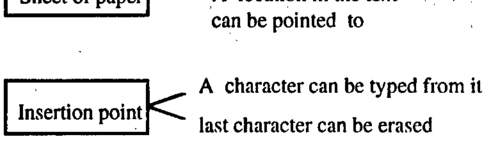

Sheet of paper
A location in the text can be pointed to
Insertion point
A character can be typed from it
last character can be erased

Figure 1. Actions that can be performed with a typewriter.

One interest in seeing analogy as a property attribution mechanism through categorization is that it does not imply that the starting point of the analogy requires a complex representation of the target. The participant might have a very crude representation of the new situation before having selected a source, as is probably the case in the situation of learning how to use a text editor, as well as in many other situations of analogy in which the new situation is really unfamiliar. This issue of how a first representation of a target situation is built have been questioned in several recent works (Bassok & Olseth, 1995; Hofstadter, 1995; Ross & Bradshaw, 1994). In our view analogy is involved in the first encoding of the new situation because the category selected provides both means of action (such as typing on a key) and means of encoding of the situation. For instance, if the task is, with a text editor, to turn 'ana logy' into 'analogy' and if the participant has no knowledge concerning how to delete a space or how to move a string of characters, as it is the case if the typewriter is taken as a source (Figure 1), the only encoding available for a true novice is that 'logy' must be deleted and written again after 'ana'. If one knows how to move a string of characters, she or he can code the task as: the part of the words have to be put closer; which leads to the cut-and-paste procedure. If the participant knows that the space can be deleted, for instance knowing already that the space is a kind of character, the situation can be coded by deleting the space and using an associated procedure like dragging then clearing, or using the backspace key. The last two codings imply (Sander & Richard, 1997) that other than the typewriter sources have been selected. In these cases, the way the situation is coded depends on the source which has been selected and thus can not be seen as the entry to the analogy mechanism.

Application to classical situations of analogy making

Gentner's example of the electric battery as like a reservoir (Gentner, 1983) is relevant in this context. It can be argued that a reservoir is defined in first place by its functional property, that is by the goal that it permits to achieve. This property is to permit storage for a differed utilisation. Thus, a reservoir is a member of the category of objects that permit storage for a differed utilisation. Following Glucksberg & Keysar (1990), it can be considered that the name of the category is one of a typical member, in which case an electric battery is an instance and reservoir is the name of the category (Figure 2). Electric battery is considered as a member of this category and the property permitting storage for a differed utilisation is attributed to it. Thus, it can be said that an electric battery is a reservoir. If we call reservoir1, a reservoir made of metal and hollow, which permits to store liquid, and reservoir2 the category of the objects which permit storage for a differed utilisation, an electric battery is a reservoir2.

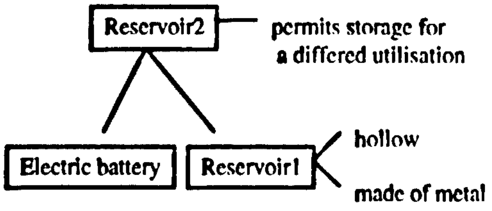


Figure 2. Analogy between an electric battery and a reservoir.

Analogy between Aesop's «sour grapes» fable and Harry's story. Consider the Aesop's «sour grapes» fable as a source story, from Wharton et al. (1994, p. 67): «A fox wanted some grapes, but couldn't reach them, so he announced to his friends that the grapes were sour anyway» and Harry's story: «Harry hoped to get a new position of marketing manager, but was passed over, so he told his wife the job would have been boring» (Ibid.).

We consider that if the source situation is known and understood by the participant, it implies that, while reading Aesop's fable, he or she will have build the category (or will have added to it, if it already exists) of situations in which a thing that is desired cannot be obtained and hence is denigrated. It is not improbable that such a category already exists, as it is usual to notice that things that we can't obtain are denigrated. An expression 'sour grapes' even exits to designate such situations. Even if this category does not exist yet for the participant, she or he will construct it while understanding the story. Aesop's fable will become a typical member of the category. While reading Harry's story, it will be categorized as another example of the same category (Figure 3).

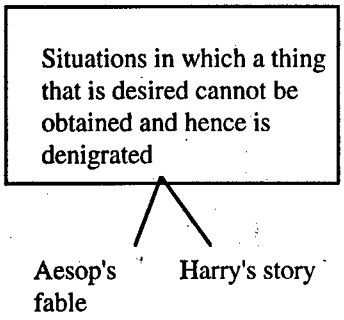

Situations in which a thing that is desired cannot be obtained and hence is denigrated
Aesop's fable
Harry's story

Figure 3. Analogy between Aesop's fable and Harry's story.

ANALOGIES INVOLVING AN ABSTRACTION PROCESS

Analogy making is not always a straightforward categorization. The first way we categorize situations can make the analogy difficult. We might build, in a first step, representations of two situations without analogical connection between them. We might also discover, after having attributed to a target the properties of a certain source, that this analogy is limiting, as it is the case when a text editor is considered as a typewriter. It is crucial to provide a mechanism explaining how this can be overcome, that is how the analogy can be discovered (or why it can not, if that is the case) or how another source, other than the first one selected, can be used if the first analogy revealed itself to be limiting.

The view that we propose for both cases is based on an ascending search in the network and can be summarized as accessing an abstract source which is more adequate.

A task which requires reaching an abstract source is difficult for at least three reasons. First, the learner has to discover a more general category to which the object may be assimilated (for instance, in the context of text editing, it is not obvious to see a digit as a manipulatable object). Second, the goals to be considered at a superordinate level are less specific (for instance, to destroy an object is a general goal which could be specified as burning it, were it to be made of wood, killing it, were it an animal, erasing it, were it a word). Third, the procedures as well are less specific or are even lacking so that the goal may be conceived but not achieved (one may have a goal without a procedure to achieve it, such as not having a procedure to destroy a piece of metal,). For these reasons, we predict that the more abstract or general the source domain relative to the target domain is, the more difficult the analogy becomes.

In the work of Sander and Richard (1997), we have shown that progress in learning was guided by analogies with sources of higher level of abstraction. We considered two categories more abstract than typewriting, ordered by an abstraction relation, namely writing in general (typewriting is a specific way of writing in general, as handwriting is another specific way); and manipulating objects (we manipulate the components of a text when we write it, when we correct it, when we duplicate and move parts of the text from one place to another). We first identified the knowledge concerning each of those categories by placing the participants in the relevant context (for instance manipulating tokens for the context of manipulating objects) and asked them to solve tasks isomorphic to the ones that can be solved on a text editor (a task of moving a string of contiguous colored tokens was isomorphic to a task of moving a word with a text-editor). Doing this with all the objects and all the goals involved allowed us to identify the knowledge about the hypothesized sources (Figure 4.a and 4.b) and to compare the learning which was actually observed with the successive use of these sources.

Once knowledge about typewriting revealed itself to be inadequate, tasks were first solved by using knowledge about writing in general (i.e., using the properties associated with the objects in

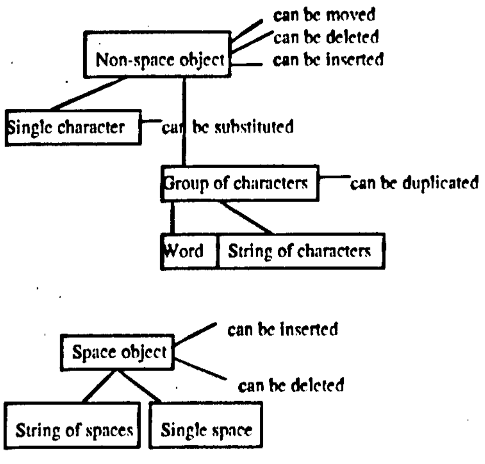

can be moved
can be deleted
can be inserted
Non-space object
Single character
can be substituted
Group of characters
can be duplicated
Word
String of characters
Space object
can be inserted
can be deleted
String of spaces
Single space

Figure 4.a. Network representing knowledge associated with writing in general and relevant for text editing.

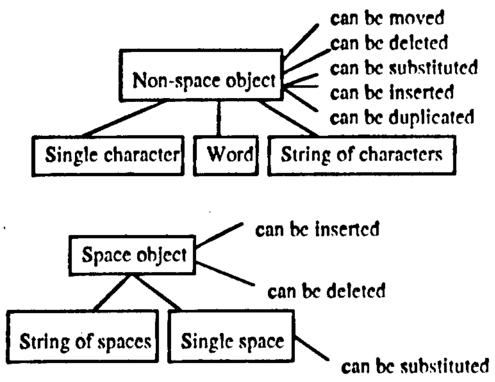

can be moved
can be deleted
can be substituted
can be inserted
can be duplicated
Non-space object
Single character
Word
String of characters
Space object
can be inserted
can be deleted
String of spaces
Single space
can be substituted

Figure 4.b. Network representing knowledge associated with manipulating objects and relevant for text editing.

the network of Figure 4.a) , or if the writing level was inadequate, knowledge about manipulating objects (i.e., using the properties associated with the objects in the network of Figure 4.b). It is in this order that participants progressively discovered the properties of the text editor. Thus, the analogy with typewriting is only the first step of learning. This stage represents the participant's entry into the semantic network and, subsequently, the entire learning process revealed to be guided by analogy with increasingly higher levels in this network. In the work that we completed on learning text editor functions, we were able to identify very precisely which semantic network was activated by the device and the tasks (knowledge represented in Figures 4a and 4b was actually tested). We will now show how data obtained in the framework of different paradigms on analogy can be analysed from the same perspective.

APPLICATION TO CLASSICAL SITUATIONS OF ANALOGY MAKING

Take, for instance, a classical problem solved by analogy, the one of Archimedes, who was asked by his king to determine whether a crown was pure gold. Because the per-volume weight of gold was known, it would have been easy to provide the answer if the volume of the crown was known. However, the crown was too ornate to measure its volume. Archimedes solved the problem while bathing. He noticed that the volume of water displaced by his body was equal to the volume of his body, so the same should hold true for the crown. In our view, the crucial point in this analogy is that the crown, as a body, is seen as having the very general property of all concrete objects, that is, in water, they displace a volume equal to their own. At the specific level, there are very few similarities between a body and a crown. Thus, the solution can not be found with that specific analogy, but at a more general level, the one of concrete objects, the relevant analogy can be drawn (Figure 5).

The crown has to be considered as a concrete object, which implies neglecting its specific properties such as symbol of kingship, made of precious metal, etc.. As well, the human body is a living body and has to be considered as a lifeless body to be put in the same category with the crown.

Gick and Holyoak (1980) consider story analogs using Duncker's (1945) radiation problem in which a tumor has to be destroyed by rays without destroying healthy tissues. In what is called the convergence solution, in which several low intensities rays are directed toward the tumor from different directions, a basic difficulty is to «think of rays as having the property of divisibility» (p. 318). In our view, a good candidate for a spontaneous analogy with rays involved in the experiment would be a ray of

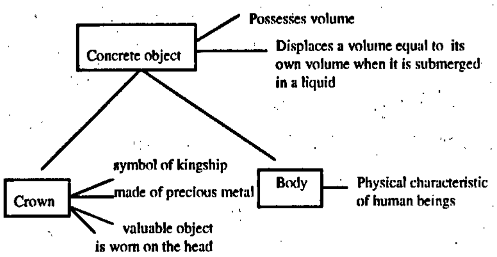

Possesses volume
Concrete object
Displaces a volume equal to its own volume when it is submerged in a liquid
symbol of kingship
made of precious metal
Crown
valuable object
is worn on the head
Body
Physical characteristic of human beings

Figure 5. Archimedes' analogy.

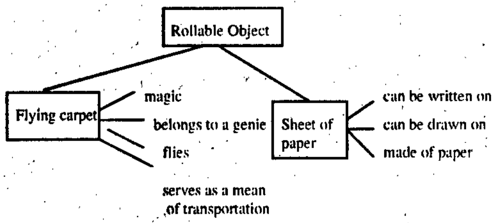


Figure 6. Analogy with the genie's story.

light: at this level, the relevant properties are «the intensity can vary» and «it can be directed in different ways». So we can predict that a large number of participants will produce solutions using these properties. Producing a convergence solution requires considering the rays as having the property of divisibility. We can predict that it will be more difficult to produce a convergence solution because the property of divisibility is not attributed to a ray of light. This prediction is supported by Duncker's (1945) results concerning participants who produced solutions without receiving solutions to analog problems: 5% spontaneously produced the convergence solution versus 29% who produced the open passage solution (of putting a tube in the esophagus), and 40% produced a kind of operation solution (of creating a tunnel in healthy tissues). The last two solutions do not require considering the property of divisibility but only the fact that rays can be freely oriented. Moreover, we can predict that if the target is changed from a ray to an object that naturally has the property of divisibility, the frequency of the convergence solution will increase because the first level reached would be the divisible-object level. Gick and Holyoak (1980) reported that Duncker found an increase in the frequency of the convergence solution when the term used was particles instead of rays. Contrary to a ray, a natural property of a group of particles is divisibility, so there is no longer a need to reach a more abstract level.

If the participants are provided with the army analog as a source (in this story, a fortress has to be captured by an army without the army being destroyed by mines), to draw the correct analogy (the convergence solution), the army has to be regarded as being composed of separable parts (soldiers or groups of soldiers) and the moving of the army must be considered as the moving of as many parts. A ray cannot be divided into parts like a solid object with unconnected parts; it divides by division of the ray sources. This requires accessing a more abstract property of division, which includes both division by separating into parts and division by dividing the source. For this reason, it can be predicted that it will be more difficult to produce the convergence solution. As a matter of fact, if participants are given a source such as the military problem in which the divisibility of the army is the relevant feature for the solution, and no hint to use this story, the convergence solution is seldom produced (Gick & Holyoak, 1980).

A study by Holyoak, Junn, and Billman (1984) also provides supporting evidence that the difficulty of the task increases with the level of abstraction required. Children had to devise as many ways as possible of transfering balls from one bowl to another. A source analog was a genie who ordered his magic carpet to roll up into a tube and then used it as a bridge to transfer jewels from one bottle to another. Among the materials provided, there was a tube and a sheet of heavy paper. According to the authors' analysis, a key factor is the «rollability» of the materials. A tube is obviously rollable, because it is already rolled, but a sheet of paper is used actually only for writing or drawing, so considering it as an object which can be rolled requires regarding it as belonging to the more general class of rollable objects and to neglect the property of being usable to write on. Thus, the mapping of a sheet of paper with the magic carpet can be done only at quite an abstract level (Figure 6) and it can be predicted that the tube solution will be easier than the sheet solution. Indeed, most of the children spontaneously used the tube, even without a story analog, but only a few participants in the story group produced the analogous rolled paper solution, even after a hint.

CONCLUSION

The main implications of our view that we wish to undeline are the following. (a) As we consider analogy as a categorization process, we are able to treat situations in which the person has a very crude representation of the target. The source participates in the encoding of the situation. (b) As our view involves an abstraction mechanism which permits to predict how the representation of the target will evolve, we can treat the issue of rerepresentation, that is how analogy can be used to deeply change a representation of a new situation and not only to add a few new relations to an existing representation. (c) We provide a formalism in which semantic (the network is a semantic network), pragmatic (goal related aspects are considered as properties of objects) and structural (the structure of the network guides the analogy mechanism) aspects are integrated in the same network. (d) As the structure of the network constrains the process, it provides a constraint system that limits combinatory explosion. (e) Semantic aspects are central in our view because semantic knowledge is used not only to decide if some objects have to be mapped but actually to infer knowledge about the new situation. This fits well with recent results (e.g., Bassok & Olseth, 1995) showing that superficial aspects of a situation are used to infer its structure.

REFERENCES

Anderson, J. R. (1991). The adaptative nature of human categorization. *Psychological Review, 98,* 409-429.

Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition, 10,* 82-93.

Barsalou, L.W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation, Vol. 27* (pp. 1-64). New-York: Academic Press.

Bassok, M., & Olseth, K.L. (1995). Object-based representations: Transfer between cases of continuous and discrete models of change. *Journal of Experimental Psychology: Learning, Memory and Cogntion, 21,* 1522-1538.

Duncker, K. (1945). On Problem Solving. *Psychological Monographs, 58,* Whole № 270.

Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive Science, 7,* 155-170.

Gentner, D. (1988). Metaphor as stucture mapping: the relational shift, *Child development, 59,* 47-59.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306-355.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: beyond similarity. *Psychological Review, 97,* 3-18.

Hofstadter, D. (1995). *Fluid concepts and creative analogies.* New York: Basic Books.

Holyoak, K. J., Junn, E. N., & Billman D. O. (1984). Development of analogical problem-solving skill. *Child Development, 55,* 2042-2055.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: The MIT press.

Kokinov, B., & Yoveva, M. Context effects on problem solving. *In Proceedings of the 18th Annual Conference of the Cognitive Science Society.* Erlbaum, Hillsdale, NJ.

Mitchell, M. (1993). *Analogy making as perception: A computer model.* Cambridge, MA: MIT Press.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological review, 92,* 289-316.

Morris, M.W., & Murphy, G.L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition, 18,* 407-418.

Poitrenaud, S. (1995). The PROCOPE semantic network: An alternative to action grammars. *International Journal of Human-Computer Studies, 42*, 31-69.

Richard, J-F., & Tijus, C.A. (1998). Modelling the affordance of objects in problem solving. *Analise Psycologia*. Special issue on cognition and context.

Rosch, E. (1978). Principles of categorization. In E. Rosch and B.B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Ross, B.H., & Bradshaw, G.L. (1994). Encoding effects of remindings, *Memory & Cognition, 22,* 591-605.

Sander, E. (1997). *Analogie et Catçgorisation*. PhD thesis, University of Paris8, Saint-Denis, France.

Sander, E., & Richard, J-F. (1997). Analogical transfer as guided by an abstraction process: The case of learning by doing in text editing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1459-1483.

Smith, E.E., & Medin, D.L. (1981). *Categories and concepts*. Cambridge: Harvard University Press.

Spalding, T. & Murphy, G.L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22*, 525-538.

Tanaka, J.W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457-482.

Turner, M. (1988). Categories and analogies. In D. H. Helman (Ed.), *Analogical Reasoning* (pp 3-24). Kluwer Academic Publishers.

Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology, 15*, 121-149.

Vosniadou, S. (1989). Analogical reasoning as a mechanism in knowledge acquisition: a developmental perspective. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 413-437). Cambridge: Cambridge University Press.

Wharton, C.M., Holyoak, K.J, Downing, P.E., Lange, T.E., Wickens, T.D., & Melz, E.R. (1994). Below the surface: Analogical similarity and retrieval competition in reminding. *Cognitive Psychology, 26,* 64-101.

Wisniewski, E.J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 21*, 449-468.

WITTGENSTEIN AND THE ONTOLOGICAL STATUS OF ANALOGY

Michael Ramscar
Department of Artificial Intelligence University of Edinburgh M.J.A.Ramscar@ed.ac.uk

Ulrike Hahn
Department of Psychology University of Warwick U.Hahn@warwick.ac.uk

ABSTRACT

Analogy has traditionally been defined in terms of a contrast definition: analogies represent connections between things which are distinct from the 'normal' connections determined by our 'ordinary' concepts and categories. A similar state of affairs holds in the case of metaphor. In order for definitions such as this to carry weight, an account of what constitutes an association between two things such that they are members of the same category rather than different ones is needed. In this paper, we explore the possibility that categorisation research might not be able to formulate a story about categories that yields the kind of unitary theoretical account that definitions of analogy and metaphor would seem to require. In particular, we focus on Wittgenstein's analysis of concepts and categories in the Philosophical Investigations (1953), and the challenges this analysis presents for contemporary accounts of categorisation. We then look at how far current accounts of categorisation can go towards meeting these challenges, and in the light of this, we evaluate the kind of ontological status that analogy (and by extension metaphor) should be given in studies of cognition. Should analogy and metaphor be seen as a separate process, definable in contrast to categorisation, or should analogy, metaphor and categorisation instead all be viewed in a wider context, as manifestations of the same underlying process?

INTRODUCTION

The belief that analogy and categorisation are distinct and separable cognitive processes is widespread: in the pursuit of our everyday lives we accept without question an ontology that distinguishes between literality - saying what something 'really' is - and analogies and metaphors, which, however informative they may be, are nevertheless not considered to make 'real' statements about the world. We may talk of "the foundations of a theory"; we may wish to "buttress a theory with more facts"; we may accept that "theories we construct can also collapse", but from our everyday viewpoints, an igloo and a castle and a skyscraper appear to share a real relationship that buildings and theories do not. We can talk of someone's foxy cunning without really meaning to imply a direct equation between the cognition of foxes and humans when it comes to being cunning. French (1995) describes how his suggestion - to an academic audience - that an upturned orange-crate, when covered with a cloth and laid out with a picnic, might really be described as a table met with the uncompromising response, "An orange crate is an orange crate is an orange crate!" The attachment to pre-theoretical intuitions is a strong one, even amongst those who seek to explore and explain them.

Research into categorisation, analogy and metaphor has usually tacitly accepted this realism. Holyoak and Thagard (1995) describe a world in which "we think we see things as they really are", and analogy is used in order to re-cycle our existing knowledge of the real world to formulate new bits of 'real' knowledge. In the literature, analogy is consistently defined in contrast to categorisation (Clement and Gentner, 1991; Holyoak and Thagard, 1995), for example, Holyoak and Thagard (1995, p217) describe analogy and metaphor as things that connect "two domains in a way that goes beyond our normal category structure".

In order to make a contrast definition stick, one needs an account of at least one of the contrasting elements. Thus, when an analogy is defined as an associative judgement between two things that are in different categories, what is needed is an account of what constitutes an association between two things such that they are members of the same category rather than different ones. In this paper, we explore the possibility that categorisation research might not be able to formulate a story about categories that yields the kind of unitary theoretical account that definitions of analogy would seem to require. In particular, we focus on Wittgenstein's analysis of concepts and categories in the Philosophical Investigations (1953; PI), and the challenges this analysis presents for contemporary accounts of categorisation. We shall then look at how far current accounts of categorisation can go towards meeting these challenges, and in the light of this, we shall evaluate the kind of ontological status that analogy (and by extension metaphor) should be given in the cognitive pantheon. We shall argue rather than viewing analogy as a separate process, definable in contrast to categorisation, both analogy and categorisation might better be seen in a wider context, as manifestations of the same underlying process.

Wittgenstein and categorisation

Previously (Ramscar, 1997; Ramscar & Hahn, 1998) we have examined in detail the veracity of the interpretation of Wittgenstein's view that is commonly held by researchers studying categorisation, comparing it with a detailed exposition of Wittgenstein's arguments. Although Wittgenstein is often presented as an opaque, difficult to interpret, and rather obscure philosopher - sometimes leading to the Philosophical Investigations being seen as a philosophical pick 'n' mix, a series of gnomic quotables to be plundered in support of a thesis - we have argued that PI sections §66 to §82 actually lay out a clear, if intricately connected, series of arguments detailing Wittgenstein's theoretical treatment of categories and categorisation in a fairly straightforward manner.

The picture that emerges from a close reading of Wittgenstein's text is at considerable variance with the general understanding of Wittgenstein's position within cognitive science, a nicely summarised account of which is presented by Lakoff (1987a; accounts which concur broadly with this can be found in Johnson-Laird, 1983; Medin & Ortony, 1989; Komatsu, 1992). Lakoff acknowledges Wittgenstein as the first theorist to notice what he terms a major crack in the classical theory of concepts and categories (e.g. Katz, 1972). Wittgenstein, claims Lakoff, argues that categories such as game cannot be accounted for according to classical theories because there are no properties that are common to all games. Lakoff draws two key theses from this argument:

1: "Games, like family members are similar to one another in a variety of ways"; and

2: "That [family resemblances], and not a single well defined collection of common properties is what makes game a category" (Lakoff, 1987a, pp 16-17)

Whilst 1 is an uncontentious statement of Wittgenstein's views, 2 is a rather more difficult interpretation to sustain. In PI §66 (p 31) Wittgenstein explicitly states that 'you will not see something that is common to all [games]'. Rather, he argues that what games have in common is the now notorious family resemblances: 'a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail' (PI, p 32). Lakoff, (and cognitive scientists in general) take this to be Wittgenstein's characterisation of what a category is. What seems to escape previous interpreters is the extreme negativity of this characterisation. In PI §67 (pp 31 -2) Wittgenstein explicitly condemns this characterisation of naming categories as vacuous. Saying that the common theme that runs through a category is the continual overlap of family resemblances is directly analogous to saying that the common thing that runs through a thread is continuous overlapping of the fibres that make up the thread, and Wittgenstein dismisses both of these accounts as empty gestures: 'Now you are only playing with words' (PI p 32). There is, he says, no thing that runs through a thread in the form of overlapping fibres; a thread simply is a series of overlapping fibres. His view is a serious challenge to, rather than an endorsement of, Lakoff's formulation: if family resemblances are the common thing that run through game, just as overlapping fibres are the common thing that run through a thread, then what is this thing supposed to be? How is it supposed to do whatever it is it is supposed to do? How long, Wittgenstein asks, is a piece of string?

Naming and boundaries - the length of a string

The question of 'how long is a piece of string?' becomes important once the second part of Lakoff's exposition is introduced. Wittgenstein, as Lakoff notes, argues that the boundaries of categories are not fixed, commenting

68.Wittgenstein and the Ontological Status of Analogy

"All right: the concept of number is defined for you as the logical sum of these individual interrelated concepts: cardinal numbers, rational numbers, real numbers, etc.; and in the same way the concept of a game is the logical sum of a corresponding set of sub-concepts." - It need not be so. For I can give the concept 'number' rigid limits in this way, that is use the word "number" for a rigidly limited concept, but I can also use it so that the extension of the concept is not closed by a frontier.. (Wittgenstein 1953, p32-3).

Lakoff interprets this discussion of number as follows: historically, numbers were first taken to be integers, and then 'numbers' were successively extended to include rational numbers, real numbers, complex numbers, transfinite numbers, and all of the other numbers that mathematicians are wont to invent. But the concept of 'number' is not bounded in any natural way, and it can be limited or extended depending upon one's circumstances and purposes. Lakoff says that in mathematics, intuitive human concepts like number must receive precise definitions: Wittgenstein's point, he claims, is that different mathematicians give different definitions, depending upon their goal. Thus although the category number can be given precise boundaries in many ways, 'the intuitive concept is not limited in any of those ways; rather, it is open to both limitations and extensions' (Lakoff, 1987a, pp 17).

The key question, on Lakoff's account, is how those limitations and extensions are governed - what factors determine the boundaries of categories in given circumstances. Lakoff answers this question in relation to game by saying that game's boundaries are governed by resemblance to previous games in appropriate ways: a new thing can be a game if it is suitably similar to previous games. Lakoff cites the introduction of video games in the 1970s as a recent example of the boundaries of the game category being extended on a large scale.

Again, discrepancies can be distinguished between Lakoff's characterisation of Wittgenstein's views and the content of Wittgenstein's stated arguments. In §68, Wittgenstein says that one 'can give the concept 'number' rigid limits in this way, that is use the word "number" for a rigidly limited concept,' - Lakoff's claim that in mathematics number must receive precise definitions appeals to this - 'but I can also use it so that the extension of the concept is not closed by a frontier.' Here, Wittgenstein is not talking about the extensibility of borders, but something far more radical: 'You can draw [a boundary], for none has so far been drawn. (But that never troubled you when you used the word "game" before)' (PI pp 32-3). Wittgenstein isn't talking here about the extensibility of boundaries; he is talking about their absence, a point developed in PI §69 to §73: categories do not have, or need, boundaries at all. In the context of Wittgenstein's overall discussion of categories, this is a vitally important point: it is one thing to seek to determine the length of a piece of string whose length is not fixed (we might add a temporal dimension to our answer for instance); it is quite another thing to seek to find out how long a piece of string is when the string is of no particular length at all.

Here, Wittgenstein is emphatic (PI §69): one can draw a boundary, for a special purpose, but it is just that, a drawn boundary. Important in the context of the special purpose, no doubt, but arbitrary to the concept or category in question. We do not need to draw boundaries, because we can happily use concepts where no boundary has been drawn; thus categories do not need boundaries to be usable. To further iterate this point, Wittgenstein considers the state of a user of a category (concept) who cannot specify that category's boundaries: is the user ignorant of those boundaries? - No, she does not 'know the boundaries because none have been drawn' (PI, p33). Not knowing the boundaries of game is not a state of ignorance - it is just reflective of the boundariless state of the category game.

> 71.*One might say that the concept 'game' is a concept with blurred edges. - "But is a blurred concept a concept at all?" - Is an indistinct photograph a picture of a person at all? Is it even always an advantage to replace an indistinct picture by a sharp one? Isn't the indistinct one often exactly what we need?*
>
> *Frege compares a concept to an area and says that an area without boundaries cannot be called an area at all. This presumably means that we cannot do anything with it. - But is it senseless to say: "Stand roughly there"? Suppose that I were standing with someone in a city square and said that. As I say it I do not draw any kind of boundary, but perhaps point with my hand - as if I were indicating a particular spot. And this is just how one might explain to someone what a game is. One gives examples and intends them to be taken in a particular way. - I do not, however, mean by this he is supposed to see in those examples that common thing that I - for some reason - was unable to express; but that he is now going to employ those examples in a particular way. Here, giving examples is not an indirect means of explaining - in default of a better. For any general definition can be misunderstood too. The point is that this is how we play the game. (I mean the language game with the word "game".) (Wittgenstein 1953, p34).*

Wittgenstein's rejection of boundaries - and not just the idea of fixing upon this boundary rather than that one - seems to be both clear and unambiguous. We don't have to define boundaries in order to use concepts, nor is it clear that definite boundaries are always what we need; these points can be further drawn out if we contemplate §71 in conjunction with §76:

> 76. *If someone were to draw a sharp boundary I could not acknowledge it as the one that I too always wanted to draw, or had drawn in my mind. For I did not want to draw one at all. His concept can be said to be not the same as mine, but akin to it. The kinship is that of two pictures, one of which consists of colour patches with vague contours, and the other of patches similarly shaped and distributed, but with clear contours. The kinship is just as undeniable as the difference. (Wittgenstein 1953, p36).*

Categories do not have boundaries, and by defining boundaries we do not capture these categories, we create something new - call them bounded categories (in §68, Wittgenstein calls them 'rigidly limited' concepts, so we might call a bounded game a rigidly limited game) - which have some kind of kinship with our natural naming categories (e.g. game), but a rigidly limited game is markedly and importantly different to game.

To return to family relations, these are the fibres that make up the threads that are categories: but Wittgenstein explicitly states that the length of these threads cannot be determined.

Wittgenstein argues that in explaining what game is, one gives examples of instances game, and one intends those examples to be taken in a particular way. What one does not do is expect the person to whom one is explaining 'game' to see the common thing - whether it be a core, schema or essence - which one cannot actually see oneself. It is true, says Wittgenstein, that when we give these examples our subject might see kinships between the examples, but these kinships are not in any way essential. Giving these examples, says Wittgenstein, is not an indirect explanation; it is the explanation. We don't give a general definition, but this is not because we can't think of one, but because there is none to give.

72. Seeing what is common. Suppose I show someone various multi-coloured pictures, and say: "The colour you see in all these is called 'yellow ochre'". - This is a definition, and the other will get to understand it by looking for and seeing what is common to the pictures. Then he can look at, and point to, the common thing

Compare this with a case where I show him figures of different shapes all painted the same colour, and say: "What these have in common is called 'yellow ochre'".

And compare this case: I show him samples of different shades of blue and say: "The colour that is common to all these is what I call 'blue'". (Wittgenstein 1953, p34).

It is not just that there is no single 'thing,' common to all: Wittgenstein questions the way that 'commonalities' are supposed to be garnered in the first place. In the first example in §72 above, the commonality is easy to spot: provided the only common colour in the pictures was yellow ochre, and provided that the subject had grasped the meaning of colour, then she will be able to grasp what yellow ochre is - the colour that is common in all the pictures.

In example two, the subject could not proceed in the same way: although the figures all have colour (yellow ochre) in common, they also have other commonalities, such as being figures. Thus the subject could as easily learn to apply 'yellow ochre' to yellow ochre or to figures, or even to samples (all of the samples are 'samples' after all) from this example. Nothing in the definition picks out the particular commonality that 'yellow ochre' is supposed to pick out[1]

Finally, in example three, there is no a priori colour commonality to the pictures; rather, the commonality can only be perceived if one already has the concept 'blue' (Otherwise, one would see a riot of various 'colours'; since understanding this example is dependent upon an understanding of 'blue', the example could not serve as an explanation of, or a definition of 'blue'.

Wittgenstein poses a number of questions that the introduction of the idea of a generalised schema to serve as the basis for a category poses. Firstly, there is the question of the form that a generalisation should take: i.e. what shape should a generalised leaf be? Linked to this is the question of the use of schemas. Even when we can answer the first question - how we say generate a generalised temperature for ice-cream - we are still left with the related question of how such a generalisation is to be used. Which particular aspects of the schema are general, and which are not (we might rephrase this question as asking which parts of the schema represent 'the generalised concept', and which are implementational details of the representation of this generalisation), and how in use are we supposed to know which is which. Is the generalised green shape a schema for green or a schema for generalised shape. Which raises the further question: provided one could generate answers to these very challenging questions, what is supposed to be intrinsic to such a schema that would cause it to be used differently to an example of that which it was supposed to be a generalisation of? In the Philosophical Investigations Wittgenstein makes quite clear his belief that no satisfactory answers to these questions can be provided. Thus he does not advocate schemas as a theory of category representation (as argued by Johnson-Laird, 1983), but rather he seeks to demonstrate that schemas alone cannot provide an account of how concepts are represented

---

[1] Quine (1960). makes a similar point in his famous gavagai discussion.

Wittgenstein and 'family resemblances'

We can state the broad outline of Wittgenstein's arguments as follows:

1. That categories have no necessary or sufficient defining characteristics: rather that kinships "family resemblances" can be traced across categories (§65-7)
2. That these category spaces are unbounded - i.e. there are no boundaries to the space across which "family resemblances" can be traced (§68, 69, 70, 71, 73)
3. That learning a category such as game does not involve extracting an essence or schema from instances. (§71-83) Rather, this process involves learning examples (instances) and appropriate ways of using these examples (§69,71, 73, 81, 82)

These arguments, as examined so far, do not advocate a particular view of concepts and categories - what has become known loosely as 'family resemblance theory' - but rather they represent a thorough attempt to elucidate the deep problems inherent in trying to account for concepts and categorisation. To Wittgenstein, the problems involved in explaining how categories are defined stem not from the phenomenon under examination, but the way this phenomenon has traditionally been defined (hence, perhaps, the famous 'don't think, but look!'). If we 'think' - i.e. if we assume that the existence of things called games entails the existence of, say a central schema (defined in some as yet to be determined way) in virtue of which the things can be considered games - we do not explore categorisation: we merely predetermine the explanations we can formulate.

Empirical support

Each of the main claims Wittgenstein makes are, we think, amply supported in the categorisation literature (for a full review of this, see Ramscar and Hahn (forthcoming).

1. Necessary and sufficient conditions. Wittgenstein's first argument attacks the definitional or "Classical'" view of concepts (Smith and Medin, 1981): this holds that concepts possess definitions specifying features necessary and sufficient for the concept. This definition is the summary description of the entire class used in every instance of categorisation, which proceeds simply by checking for the presence of these features in the entity in question. This view is commonly supplemented by the "nesting assumption" that a subordinate concept (e.g.. robin) contains nested within in it the defining features of the super-ordinate (bird).

However, the definitional view seems inadequate as a theory when transferred from artificial concepts in controlled experiments to our everyday concepts (i.e. the concepts for which we typically have words). Of the difficulties faced here, the most serious one is that almost all everyday concepts appear to be indefinable (Fodor, 1981). It simply does not seem possible to formulate necessary and sufficient conditions for being, for example, a chair, or a window, or a smile; illustrated by the fact that dictionary "definitions" of almost all terms are not really definitions at all. They do not provide necessary and sufficient conditions for category membership - instead they typically do no more than provide some relevant information about category members, which may help the dictionary user identify which concept in intended. Further evidence against the definitional view comes from examining the boundaries of natural language categories. The definitional view implies that these are sharp, cleanly separating instances from non-instances. But, as Wittgenstein claimed, this turns out not to be the case.

2. Boundaries. In 1949, Black provided the following thought experiment to illustrate that category boundaries might be vague: on is to imagine a series of 'chairs' differing in quality by least noticeable amounts. This can give rise to an ordered sequence which moves from a Chippendale chair on the one end to a small nondescript lump of wood at the other end. A 'normal' observer, argues Black, should find it extremely difficult to point to the dividing line between 'chairs' and 'non-chairs' along this continuum, which illustrates a different source for category vagueness. (The difficulties posed

by continua were already recognized in the sorites (heap) and and phalakro (bald man) paradoxes, which originate with the Megarian philosophers in the early 4th century, Barnes, 1979.) Black makes it clear that this uncertainty over category boundaries can be generated for any term whose application requires the use of a sense, that is to say all 'material' terms.

Quine (1960) points out that indeterminacy can arise not only because the category boundary is vague (a phenomenon generally referred to as 'fuzziness') but also because the boundaries of an entity can be vague. To illustrate with his example of 'mountain' which is "vague on the score of how much terrain to reckon into each of the indisputable mountains, and it is vague on the score of what lesser eminences count as mountains" (Quine, 1960, p. 126)

A third source of uncertainty over boundaries has been identified by Lakoff (1987b). Even when concepts do appear to have definitions, these definitions generally hold only with respect to a range of 'background assumptions'. Varying these assumptions immediately produces unclear or borderline cases:

"The noun bachelor can be defined as an unmarried adult man, but the noun clearly exists as a motivated device for categorizing people only in the context of a human society in which certain expectations about marriage and marriageable age obtain. Male participants in long-term unmarried couplings would not ordinarily be described as bachelors; a boy abandoned in the jungle and grown to maturity away from contact with human society would not be called a bachelor." (Fillmore, quoted in Lakoff, 1987b)

Background factors, such as the social conventions concerning marriage, will, in general, hold to varying degrees. Presumably the definition of bachelor can meaningfully be applied if the background conditions are sufficiently similar to the conventions concerning marriage current in the West.

Alongside such arguments, direct empirical evidence that (many) natural language categories do not have clear boundaries was accumulated in the 1970's. The first studies we know of were conducted by the linguist William Labov, summarised in Labov (1973). His empirical work focuses on cup-like containers, examining the variability inherent in the use of terms such as cup, bowl, mug etc., between subjects and between contexts. Labov's interest was primarily in formalising the variability found, thus his results are not presented with the detail experimental psychologists might want. This gap is readily filled by McCloskey and Glucksberg (1978) who presented a study of 540 exemplar-category pairs (e.g., apple-fruit) which revealed not only substantial between and within subject disagreement over category membership (the latter measured over successive test-sessions) but also showed levels of disagreement to correlate with independently derived typicality ratings.

3. Essences versus examples. Wittgenstein's final point rejects the idea of some abstracted schema in preference for an account based on previously encountered examples. Whilst, as Komatsu (1992) notes, the vast majority of experimental results do not directly indicate anything about conceptual representation: separating form, content and the processes acting on concepts is an invidious business (best illustrated by Wittgenstein's remarks on schemas above) the issue of whether or not a particular learning process involves the abstraction some core essence - be it a schema, a theory or a prototype - or not has been central to experimental psychology in the last decades and has been pursued not only in concept learning tasks, but also in related domains such as Artificial Grammar Learning (Shanks and St John, 1994). Controversy has raged not only over actual empirical evidence for or against abstraction, but also about the very criteria on which a distinction could conceptually and empirically based.[2]

---

[2] Barsalou (1990) has argued that exemplar storage and abstraction in category representation are impossible to distinguish in principle. This position, based on a highly idiosyncratic notion of abstraction, is overly pessimistic. Careful evaluation of the many criteria that have been put forth, particularly in order to distinguish between processes based on rules and processes based on exemplar similarity, reveals that many have been overestimated in their power to cleanly distinguish between the two (Hahn and Chater, 1998).

From an experimental perspective, Hahn and Chater (1998) argue, a compelling way to address this issue is through model comparisons of fully specified cognitive models. Take for example, the evidence regarding prototypes: evidence for prototypes in natural language categories has been sought from a variety of sources. Classic are those studies which identified a variety of so-called "prototype effects"; all of these involve some form of differential reaction to central or typical members of a category such as differences in typicality ratings, faster reaction times in speeded classification tasks or differential retention in memory relative to other items (see e.g. Rosch, Simpson and Miller, 1976; Posner and Keele, 1968; Posner and Keele, 1970). However, such effects do not unequivocally indicate mental representations of concepts in terms of prototypes (Lakoff, 1987b). Rather such effects might arise from cognitive representations and processes which make no use of representations of prototypes or central tendencies as such.

This is made clear by comparative modelfitting of fully specified process models (though this tends to come at the expense of artificial stimulus domains). The categorisation literature has accumulated a wealth of studies in which model comparisons between exemplar models which simply store all encountered instances in memory, and prototype models which abstract a central tendency have consistently gone in favour of exemplar models: exemplar models have yielded quantitative fits superior to the prototype models tested and accounted for a wide range of phenomena traditionally associated with prototypes such as the instability of instance retrieval and typicality judgements; the levels of specificity at which concepts are encoded; sensitivity to correlations amongst category instances; and the way accuracy in classification tasks increases with category size (Nosofsky, 1986, 1987, 1988b, 1989, 1991b, Nosofsky, Clark and Chin, 1989, Shin and Nosofsky, 1992; Lamberts, 1996). Moreover, as Komatsu (1992) notes, if one assumes that individuals only retrieve a subset of these stored exemplars on any given occasion, but are inclined to regard that subset as exhaustive (Nickerson, 1981), then the an exemplar based approach may also be able to begin to explain why it is that people believe that categories have essences and boundaries.[3]

Similarly, those few empirical studies have directly addressed the assumptions behind core essences - whether as schemas or theories - have found little or no support for the idea that essences are extracted in category learning. Malt (1994) found the assumption (Putnam, 1975) that H_2O is the essence of water did not stand up to empirical scrutiny, and that judgements of the amount of H2O in a liquid were very poor predictors of whether it was water or not. In another study, Ramscar, Darrington, Pain and Lee (1998) used differences in the recall characteristics of surface and structural aspects of representations (Gentner, Ratterman and Forbus, 1993) to show that subjects could classify items together under a category name, and carry out recall tasks with category members grouped by that name, without extracting a category schema or essence; Ramscar et al's subjects appeared to have stored only exemplars in their category encoding.

In summary, at present at least, there is no clear evidence in the literature for abstraction in concepts acquisition, whilst there is considerable evidence which can be marshalled support of some kind exemplar based account .

WHITHER TWO PROCESSES?

Like the empirical finding we present above, Wittgenstein's arguments bear down on any all-encompassing view of category structure. Together, the two appear to effectively explode the idea of the category as a unitary theoretical instrument: how likely is it that, even if categories aren't defining features, shared essences or some other common thread running through, that there is a fundamental unity in all categories? That clear cut members all have higher within category similarity than between category similarity or that all are based on partial theories, and so on?

We argued earlier that in order for the standard contrast definition of analogy to do its work, an account of categorisation, distinct from that contrast definition, was necessary. As this brief survey shows, no such account is available, nor, does it seem likely that any answers to Wittgenstein's deep questions regarding any 'straightforward' account of categorisation will be forthcoming.

Furthermore, we have carried out a number of studies which directly explore the contrast definition from the opposite direction, examining the properties typically used to separate analogy from categorisation. Ramscar and Pain (1996), showed that subjects would categorise Gentner, Ratterman and Forbus's (1993) classic analogy materials using exactly the same process that they used to determine analogies between them. Darrington, Lingstadt and Ramscar (1998) showed that the same process - structure mapping, typically considered the preserve of analogy - could cause subjects to override supposedly ecological categories in sorting tasks, with participants preferring groupings between pots and walls, and walls and pans to pots and pans and walls alone. These studies can be added to other theoretical and empirical evidence against a two-process account of literal (categorical) versus non-literal (analogical or metaphorical) reasoning, such as Hoffman and Kemper's (1987) review of a number of reaction time studies which also demonstrates the paucity of the evidence for the widely held belief that literal (intra-categorical) meanings are processed faster than metaphorical (intercategorical) meanings (as well as the considerable evidence for the opposite effect; see also Récanati, 1995, Glucksburg and Keysar, 1990, Gibbs, 1984). Theoretically, at least, distinguishing analogy from categorisation may not be the simple task our intuitions - and the literature - might have us believe.

One defence, in the light of these arguments, might be an appeal to categories grounded in ecology: the difference between analogy and categorisation is that categories really do - in some way - reflect the underlying structure of the world in a way that analogies do not. Whilst researchers in mainstream categorisation research are at often pains to disavow metaphysical realism (c.f. Murphy, 1996) in practice, the very kinds of categories they choose to examine, and the attitude they adopt towards them in discussing between-category comparisons, tempers the impact of these protests.

In disagreeing with Wittgenstein's position regarding categorisation, Medin and Ortony (1989) suggest that if people really think about the fact that whales are mammals not fish, they will see that with respect to some important, although less accessible property or properties whales are similar to other mammals. "If one cannot appeal to hidden properties, it is difficult to explain the fact that people might recognise such similarities... there might be a price to pay for looking rather than thinking." (Me-

---

[3] One other contender in current debate about conceptual structure is the so-called theory-based view (Murphy & Medin, 1985; Medin & Ortony, 1989). The theory-based view is defined primarily in contrast to any account, prototype- or exemplar-based, which seeks to ground real world categories in terms of perceptual similarity. It emphasises the role of background knowledge or "theories" in our everyday classification, in order to explain, for instance, the fact that, despite strong perceptual similarities, we do not classify bats as birds. Due to its lack of explicitness the theory-based view is not that easy to align with Wittgenstein's claims. Given the problems inherent in definitional accounts of conceptual structure (see above), one must assume that "theories" are not complete, i.e. they allow deduction of classification decisions, but are only "partial", in that they form one component of a complex, non-deductive overall process (Hahn & Chater, 1997). This overall process is not generally spelled out by advocates of the theory-based view. The simple claim then that "partial theories" or background knowledge are relevant to categorisation need not conflict with Wittgenstein's arguments. There is no statement about boundedness, nor is there a claim of definitional features. Though the theory-based view does suggest that learning and understanding a category also involves acquiring appropriate background knowledge, this does not directly contradict the role of examples in acquisition and use, but merely suggests an additional factor. This still leaves a problem regarding partial theories, i.e. how partial does a theory have to be to not be stating an "essence"? Given that the theory-based view has done little to provide full accounts of any categories, no definate answer can be given to this question here. To the extent though, that too much faith is invested in the power of theories, another look at Wittgenstein's arguments and examples might be sobering

din and Ortony, 1989, pp 179 - 180). The problem is that it is just this fact, that is the point of any investigation of human categorisation (c.f. Malt, 1994). In 'Ontology', (Moby Dick, Melville, 1851), the central character, Ishmael, examines all of the reasons put forward by Linnaeus for classifying whales as mammals.

> I submitted all [these] to my friends Simeon Macey and Charlie Coffin, of Nantucket, both messmates of mine in a certain voyage, and they united in the opinion that the reasons set forth were altogether insufficient. Charlie profanely hinted that they were humbug.
>
> Be it known that, waiving all argument, I take the good old fashioned ground that the whale is a fish, and call upon holy Jonah to back me

As Wittgenstein famously remarked, our talk of process and states is just what commits us to a particular way of looking at a matter, (Wittgenstein 1953, p102). Choosing what is to count as facts when it comes to categorisation is a powerful determinant of the picture of the process one will uncover. And taking on board a different set of facts can radically alter any such picture. All classification systems are human constructs, and our immersion in one such system shouldn't blind us to alternatives. Similarly, it is important to be aware of the social dimensions of categorisation, and the way collective and individual categories can differ; it may be - in the study of the cognitive processes of categorisation - that individual facts might reveal more than collective ones.

If we broaden our view, we see that ecologically, the distinction between categorisation and analogy is a recent one: the conceptual revolution begun by Linnaeus represents the overthrow by a system based on heredity of a previous system based far more on analogy. As Thomas (1984) argues in his detailed account of changes in natural kind categories in England in the period 1500 - 1800, for much of the early modern period, 'the universal belief in analogy' resulted in much of the natural world being categorised and understood by analogy with human social structures. Bees had Princes, Potentates, Kingdoms and Dominions (Warder, 1716; Rusden, 1679, quoted in Thomas, 1984 p. 62); they were ruled over by 'a fair and stately bee, having a majestic gait and aspect' (Levett, 1634, quoted in Thomas, 1984, p. 62). Cranes followed a captain; Rooks had a parliament; Storks and Ants and Beavers were avowed republicans. As Thomas notes, this picture of the natural world fed back recursively into concepts of human society: King Henry VII once ordered the execution of all mastiffs, after they had baited a lion, 'being deeply displeased ... that an ill-favoured rascal cur should with such violent villainy assault the valiant lion, king of all beasts' (Caius, 1576, quoted in Thomas, 1984, p. 60)).

The important issue here is not whether the Linnaean way of construing the world is right, or whether other 'pre- Linnaean' conceptual schemes are wrong; nor is it a question of finding an analysis that will answer these questions. All that different conceptual schemes such as these reflect is the differing attitudes to pre-theoretical ideas of categorisation and analogy that they embody (and, as Lakoff, 1987a, illustrates, the Linnaean revolution may be less complete than we generally believe). Our claim is that if we wish to explain the cognitive processes that actually underpin analogy and categorisation, then it is just these pre-theoretical intuitions we should question, and, for certain purposes, abandon.

The consequence of our investigation, of both Wittgenstein's position and the supporting evidence, is a claim analogous to that which has been made for the related process of processes that determine literal and metaphoric meaning. Gibbs (1984) notes that the claim that there is no principled distinction between literal and metaphoric meaning leaves one important question unanswered: how can we explain why people can often judge a sentence to be literal or metaphoric? What lies behind the intuition that "an orange crate is an orange crate is an orange crate"? Whilst Gibbs acknowledges that this intuition needs exploring, he asks "does it indicate that listeners **process** [our emphasis] so called literal and metaphoric utterances differently?" (p. 296). Rumelhart (1979) makes the point that "the classification of an utterance as to whether it involves literal or metaphoric meanings is analogous to our judgement as to whether a bit of language is formal or informal. It is a judgement that can be reliably made, but not one which signals fundamentally different comprehension processes" (p. 79). Gibbs argues that one reason why some sentences seem so literal is that listeners are influenced by the interpretative context in which such judgements are made: people judge a sentence as having literal meaning because it is isomorphic with the situation in which the sentence is interpreted (Fish, 1980). However, it doesn't follow from this that the literal meanings of sentences can be uniquely determined, as our understandings of situations always influence our understandings of sentences. Says Gibbs, "To speak of a sentence's literal meaning is to already have read it in the light of some purpose, to have engaged in an interpretation. What often appears to be the literal meaning of a sentence is just an occasion-specific meaning where context is so widely shared that there doesn't seem to be a context at all." Gibbs, 1984, p. 296; As for judging sentences are literal, we claim, so for judging whether whales are mammals or fish; or, for that matter, whether our picnic is 'on the orange-crate' or 'on the table'.

It may be that the best accounts of categorisation will also incorporate an account of analogy, and explain both in terms of a single cognitive process. Some of the more important findings from existing analogy research are the important role that representational structure has to play in similarity judgements, and the differing roles that surface and structural features play in recall. It may be that incorporating a dimension of structural similarity into the similarity space mapped in an exemplar model of categorisation might also enable the modelling of analogy and superficial similarity, without recourse to multiple processes. On such a model, strong similarity across all dimensions (including both surface and structural similarities) might betoken strong categorical similarity - with, perhaps, the strongest similarities occurring in basic level categories - whereas strong mappings on only a subset of similarity dimensions would underpin analogical (or superficial, or metaphorical) similarity.

This would still leave us with the problem of explaining peoples' intuitions about analogy and categorisation. However, as we noted earlier, if one assumes that individuals only retrieve a subset of stored exemplars during any given similarity computation episode, and that they may be inclined to regard that subset as exhaustive (mimicking Gibb's, 1994, point made earlier: all judgements are contextual, even if it doesn't feel like they are; the subset of exemplars recalled simply matches the context of the categorisation judgement to be made) then an exemplar based approach might be able to begin to explain why it is that people believe that categories have essences and boundaries. To return to French's (1995) suggestion that an orange-crate, when covered with a cloth and laid out with a picnic, might really be a table: a model such as this might be able to explain more than why it is that 'an orange crate is an orange crate, can be a table'. If we could show how 'ordinary' categorical judgements of table are just those occasion-specific judgements where context is so widely shared that there doesn't seem to be any context at all, we might also be able to offer an explanation of why it is that some people find this idea so very counter-intuitive.

REFERENCES

Barnes, J (1979) The Presocratic Philosophers, Routledge, London

Barsalou, L. (1990) On the indistinguishability of exemplar memory and abstraction in category representation. In: Srull, T.K., Wyers, R.S (eds) Advances in Social Cognition, Voll. III, Content and Process Specificity in the Effects of Prior Experiences. Hillsdale, NJ: Erlbaum.

Black, M (1946) Language and Philosophy. Ithaca: Cornell University.

Caius, J (1576) Of English Dogges

Clement C.A, and Gentner, D. (1991) Systematicity as a selection constraint in analogical mapping Cognitive Science, 15: 89-132.

Darrington, S, Lingstadt, T and Ramscar, MJA (1998) Analogy as a sub-process of categorisation. Proceedings of the 20th Annual Meeting of the Cognitive Science Society, Madison-Wisconsin, Earlbaum, NJ.

Fodor, JA (1981) The present status of the innateness controversy in J Fodor (ed) Representations, MIT Press

Fish, S (1980) Normal circumstances, literal language, direct speech acts, the ordinary, the everyday, the obvious what goes without saying, and other special cases. In S Fish (ed) Is there a text in this class? Harvard University Press

French, RM (1995) The Subtlety of Sameness, MIT Press.

Gentner, D Ratterman, M.J. and Forbus, K. (1993) The roles of similarity in transfer. Cognitive Psychology **25**: 524-575

Gibbs, RW (1984) Literal meaning and psychological theory. Cognitive Science 8:275-304

Glucksberg, S. and Keysar, B. (1990) Understanding metaphorical comparisons: Beyond Similarity. Psychological Review, **97**: 1: 3-18

Hahn, U. & Chater, N. (1997) Concepts and Similarity. chapter in: Knowledge, Concepts and Categories, pp. 43-92, K. Lamberts & D. Shanks (eds.), MIT Press.

Hahn, U and Chater, N. (1998) Similarity and Rules: Distinct? Exhaustive? Empirically Distinguishable? Cognition, 65, 197-203.

Hoffman, R R and Kemper, S (1987) What could reaction-time studies be telling us about metaphor comprehension? Metaphor and Symbolic Activity **2(3)**, 149 - 186

Holyoak, K.J. and Thagard, P. (1995) Mental Leaps. MIT Press.

Johnson-Laird, P.N. (1983) Mental Models. Cambridge University Press, Cambridge.

Katz, JJ (1972) Semantic Theory, Harper & Row, New York.

Komatsu, L K (1992) Recent views of conceptual structure. Psychological Bulletin, **112**(3), 500-526

Labov, W. (1973) The boundaries of words and their meanings. In, Bailey, C-J.N. and Shuy, R.W. (eds) New ways of analysing variation in English. Washington, DC: Georgetown University Press, pp 340-373.

Lakoff, G (1987a) Women, Fire and Dangerous Things. University of Chicago Press, Chicago, Illinois

Lakoff, G. (1987b) Cognitive Models and Prototype Theory. In, Neisser, U. (ed) Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization. Cambridge, England: Cambridge University Press.

Lamberts, K. (1996)Exemplar models and prototype effects in similarity-based categorization. Journal Of Experimental Psychology-Learning Memory And Cognition, 22, 1503-1507.

Levett, J. (1634) The Ordering of Bees

McCloskey, M.E. and Glucksberg, S. (1978) Natural Categories: Well Defined or Fuzzy Sets? Memory and Cognition, 6, 462-472.

Malt B.C. (1994) Water is not H2O. Cognitive Psychology 27:41-70

Medin, D & Ortony, A. (1989) What is psychological essentialism? In S. Vosniardou and A. Ortony (Eds) Similarity and analogical reasoning. Cambridge University Press.

Melville, H (1851) Moby Dick; or, The Whale. Harper Bros. New York.

Murphy, GL & Medin, DL (1985) The role of theories in conceptual coherence. Psychological Review, 92, 289-316.

Murphy, GL (1996) On metaphoric representation, Cognition: **60**:173-204

Nickerson, R (1981) Motivated retrieval from archival memory. In G Bower (ed.) Nebraska symposium of motivation (28, pp. 73-119) University of Nebraska Press

Nosofsky, R.M (1986) Attention, Similarity and the Identification-Categorization Relationship, Journal of Experimental Psychology: General, 115, 39-57,

Nosofsky, R.M. (1987) Attention and Learning Processes in the Identification-Categorization Relationship. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 87-109.

Nosofsky, R.M (1988a) Exemplar-based accounts of the relations between classification, recognition, and typicality. Journal of Experimental Psychology: Learning, Memory and Cognition, 14, 700-708.

Nosofsky, R.M. (1988b) Similarity, frequency, and category representations. Journal of Experimental Psychology: Learning, Memory and Cogntion, 14, 56-65.

Nosofsky, R.M. (1989) Further tests of an exemplar-similarity approach to relating identification and categorization. Perception & Psychophysics, 45, 279-290.

Nosofsky, R.M. (1991) Tests of an exemplar model for relating perceptual classification and recognition memory. Journal of Experimental psychology: Human Perception and Performance, 17, 3-27.

Nosofsky, R.M., Clark, S.E., & Shin, H.J. (1989) Rules and exemplars in categorization, identification and cognition. Journal of Experimental Psychology: Learning, Mmeory and Cognition 15, 282-304.

Posner, M.I. & Keele, S.W. (1968) On the genesis of abstract ideas. Journal fo Experimental Psychology, 77, 353-363.

Posner, M.I. & Keele, S.W. (1970) Retention of abstract ideas. Journal of Experimental Psychology, 83, 304-308.

Putnam (1975) The meaning of 'meaning' In H Putnam (ed.) Mind, language and reality: Philosophical papers, Vol. 2 Cambridge University Press

Quine, W.V.O. (1960) Word and object. MIT Press.

Ramscar, M.J.A. (1997) Wittgenstein and the nature of psychological categories. Proceedings of SimCat 97, Department of Artificial Intelligence Conference Proceedings, University of Edinburgh, Scotland, 205-211.

Ramscar, MJA & Pain, HG (1996) Can a real distinction be made between cognitive theories of analogy and categorisation? (In Proc.) Proceedings of the 18th Annual Meeting of the Cognitive Science Society. San-Diego, LEA, NJ.

Ramscar, M.J.A. and Hahn, U (1998) What family resemblances are not. Proceedings of the 20th Annual Meeting of the Cognitive Science Society, Madison-Wisconsin, LEA, NJ

Ramscar and Hahn (forthcoming) Wittgenstein the representation of categories: what family resemblances aren't. Manuscript in preparation.

Ramscar, M.J.A., Pain, HG, Darrington, S and Lee, J (1998) Examples and generalisations: using surface versus structural recall biases to probe conceptual storage Proceedings of the 20th Annual Meeting of the Cognitive Science Society, Madison-Wisconsin, LEA, NJ

Récanati, F (1995) The alleged priority of literal interpretation Cognitive Science 19:207-273

Rosch, E.H., Simpson, C & Miller, R.S. (1976) Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 2, 491-502.

Rumelhart, D (1979) Some problems with literal meanings. in A. Ortony(ed.) Metaphor and Thought. Oxford University Press.

Rusden, A (1679) A Further Discovery of Bees

Shanks, D.R. and St John, M.F. (1994) Characteristics of dissociable human learning systems. Behavioral and Brain Sciences, 17, 367-395.

Shin, H.J. and Nosofsky, R.M. (1992) Similarity-Scaling Studies of Dot-Pattern Classification and Recognition. Journal of Experimental psychology: General, 121, 278-304.

Smith, E. & Medin, D.L. (1981) Categories and Concepts. Cambridge, MA: Harvard University Press.

Warder, J (1716) The True Amazons

Wittgenstein, L trans. Anscombe, E (1953). Philosophical Investigations Blackwell, Oxford.

Abstracts

MULTIPLICATIVE BINDING, REPRESENTATION OPERATORS, AND ANALOGY

Ross W. Gayler
Department of Psychology, The University of Melbourne
Parkville VIC 3052, AUSTRALIA
r.gayler@psych.unimelb.edu.au

ABSTRACT

Analogical inference depends on systematic substitution of the components of compositional structures. Simple systematic substitution has been achieved in a number of connectionist systems that support binding (the ability to create connectionist representations of the combination of component representations). These systems have used two types of binding operators (generically renamed here as **bind()** and **bundle()**) implemented in various ways. This paper introduces a novel implementation of the **bind()** operator. This implementation is interesting because it is removes some of the complexities of other implementations, can be efficiently implemented, and allows easy specification of queries in a way that highlights their equivalence to analogical mapping problems.

The binding operators may also be viewed as representational operators because they are used for the construction of complex, compositional representations. The specific implementation of the representation operators partially constrains the representations that may be constructed. This paper shows that some binding systems are unable to adequately represent hierarchical compositional structures. A novel family of representational operators (called **braid()**) is introduced to allow representation of nested structures. Other potential uses of the **braid()** operators are also explored.

The specific implementation of the representation operators does not completely constrain the representations which may be constructed. A system designer must also choose a representational idiom for the encoding of information. The choice of representational idiom will further constrain the relative ease of different cognitive operations. The most commonly used idiom (based on frames of role/filler bindings) limits the simultaneous representation of multiple objects. This paper proposes an alternative idiom (also based on frames) to solve this problem.

The new representational idiom highlights a previously unnoticed problem (which exists in other connectionist binding systems) with maintaining the disjointness of roles and fillers. This problem is explored and several solution approaches discussed. One interesting approach depends on a generalisation of the newly introduced **braid()** operator.

The new representational idiom suggests that cognitive operations of bottom-up and top-down object recognition should be relatively easy. These operations depend absolutely on analogical mapping in order to connect disjoint representations and drive perceptual search.

ADAPTATION OF NON-ISOMORPHIC SOURCES IN ANALOGICAL PROBLEM SOLVING

Ute Schmid
Department of Applied Computer Science, Technical University Berlin
email: schmid@cs.tu-berlin.de

We propose a computational model for analogical problem solving which especially adresses the influence of structural characteristics on adaptation and learning (Gentner , 1983). While there is strong empirical evidence that semantic and pragmatic aspects are important constraints for retrieval of source problems as well as for analogical mapping (Hummel & Holyoak, 1997), we believe it worthwhile to further investigate structural properties: It is evident that in realistic settings source problems are usually not isomorphical to target problems. But the question which kind of structural properties are necessary for succesful adaptation is seldom addressed in psychological experiments (Hummel et al., 1997) and there are no computational models dealing with structure mapping *and* adaptation and learning in the case of non isomorphical problems. For example, PUPS (Anderson & Thompson, 1989) deals with adaptation and learning, but only for problem isomorphs; LISA (Hummel et al., 1997) deals with not isomorphical problems, but gives only regard to analogical access and mapping.

Our model IPAL was developed in the context of automatic programming (Schmid and Wysotyki 1998). But we believe, that it also contains useful ideas for cognitive modelling. Problems as well as problem schemes are represented in a common format, namely as graphs or trees. Mapping between two problems (or a current problem and a problem scheme already acquired) is done by means of a tree-metric: The similarity between two structures is given by the weighted number of operations (substitution, insertion and deletion of nodes representing objects, relations or functions) needed to transform the source structure into the target. Mapping guides retrieval as well as adaptation. If two structures are isomorphical, they can be transformed into another by a unique set of substitutions. Otherwise, the source solution can be adapted to the target problem by applying the operations gained by the mapping of the problem descriptions to the solution of the source problem. If a target problem could be successfully solved by adaptation of a source, a generalized scheme, which covers the common structure of source and target, is constructed. The target problem and the generalized scheme are committed to memory with the generalized scheme as parent to source and target. Thereby, a hierarchical memory structure develops while the system gets confronted with new problems.

We have tested IPAL with a variety of structural relations between source and target pairs and obtained the following results: If source and target are isomorphical, adaptation success is 100%, for homomorphical structures (mono- or epimorphical) 66%, for problems with no defined structural relation 4%. This shows that there have to be characteristics for structural relationships not covered by the concept of morphisms. Our next aim therefore is to identify further structural constraints for adaptation success.

Additionally we have performed two experiments where the structural similarity between source and target was systematically variied. We obtained the following results: (1) people are able to adapt partial isomorphic problems (i.e. the source structure is contained completely in the target structure) only if the superficial similarity between source and target is high (Keane et al., 1994); (2) given high superficial similarity, partial isomorphs can be adapted succesfully if the number of nodes of the common structure is more or equal to the number of nodes of the (larger) structure of the target problem.

"AQuARIUM: A HIERARCHICALLY-SUPPORTED MONO-SYMBOLIC LANGUAGE FOR ANALOGIC INTEGRATION"

Ron Cottam, Willy Ranson & Roger Vounckx
Evolutionary Processing Group of the Laboratory for Mcroelectronics and Technology
VUB-IMEC Electronics Division, The University of Brussels (VUB)
Pleinlaan 2, 1050 Brussels, Belgium

We propose a new contextually aware universal paradigm which can extend or replace formal logic, and which is capable of supporting hierarchical metastates and a description of the development of life and consciousness through evolutionary computation.

In presupposing a coherent universe, we acknowledge the correlation of its constituent properties and processes, and accept that *all* of its regions must remain communicative to support coherence. Distinguishable forms then exist through the actions of *one* coherent set of processes. Successful survivalist processing of massive amounts of real-time data by living entities necessitates the availability of simplified but locally representative models of "reality" which are couched in terms familiar to the processor: the use of analogues.

The selection of favourable analogues follows the same criteria and suffers from the same difficulties as does their successful linguistic transmission. We can integrate these two processes into a single format, that of a unified hierarchical symbolic language which displays only-partially-deterministic coupling between its formally represented parts. "AQuARIUM"[1] provides a framework for this symbolic language, which consists initially of only a single symbol. The symbol contains just enough information to invite questions as to its significance, without presenting sufficient detail for an intelligently inquisitive "selector" to be sure of the correctness of an initial guess as to its meaning. The nature of the resulting questions can then be used to evaluate the context into which more detailed description will be placed, rather than presupposing unilaterally a "correct" comprehensional context.

Separate analogues emerge from "reality" as structures which correspond to the formulation of "locally sufficient" approximating metastatic representations of an otherwise partially disordered or chaotic region of the universal phase space. Consequently, an analogue is always to some extent defective in its detail, in that it must of necessity exhibit differences from its "real" counterpart. Internally, for an "originating" processor, the use of a selected analogue is relatively simple, given a good memory of which characteristics have been selected as, or determined to be, "correct" analogous details. However, the transfer of an analogue from one processor to another is fraught with dangers. The major difficulty in selecting a transferable analogue is to match the "representative" characteristics recognised by its creator to those which are interpreted by its receptor. For example, in likening the flow of "electrons" through a network of wires and switches, to the early-morning rush of commuters through tunnels and barriers in accessing the Metro, we should not assume that "electrons" carry briefcases with them, nor that first of all they kiss their wives goodbye before commencing the journey.

Communication of an *idea* from one processor to another depends on an equivalence of *both* of their logic systems *and* their data environments, or alternatively on a successful manner of evaluating any differences between these and correcting for them. This *always* necessitates a two-directional process where ultimately it will be unimportant which of the two processors initiated the communication, but only whether this evaluation and correction has been successfully carried out. The implied correspondence to inter-processor *cooperation* is inherent to the framework provided by AQuARIUM.

---

[1] AQuARIUM: "A Query and And Reflection Interaction Using MAGIC: Mathematical Algorithms Generating Interdependent Confidences".

Ultimately, in a coherent universe, *all* analogues of *all* "realities" are equivalent when account is taken of their associated approximations, and they can consequently all be integrated into a descriptive language of this kind. The maintenance of universal universal coherence requires continuous communication between all stable metastatic entities, yet the natural presence of an Einsteinian communication restriction eliminates the possibility of instantaneous direct correlation in a causally coherent domain. Formally defined metastates *cannot* communicate directly with each other, and any communication which does occur must take place *at least partially* through the causal chaos represented by nonlocality. The complete range of possibilities between these two extremes can initially be modeled in AQuARIUM by a modified recursive form of Dempster-Schafer probability.

THE MECHANISMS OF MAPPING: EVIDENCE FROM ON-LINE ANALOGY JUDGMENTS

Kenneth J. Kurtz and Dedre Gentner
Northwestern University
Department of Psychology
2029 Sheridan Road
Evanston, IL 60201-2710

An account of analogical thinking must explain structure sensitivity and flexibility in the comparison process (Gentner & Markman, 1993; Hummel & Holyoak, 1997). Analogical mapping is widely viewed as the alignment of structured representations to maximize common relational structure. The process model of structure-mapping, as operationalized in SME (Falkenhainer, Forbus & Gentner, 1989), relies on matching predicates that are identical in both the source and target. In addition, non-identical matches can be made when: 1) systems of identity-matches license correspondence between certain non-identical elements, and/or 2) semantically similar, but non-identical, predicates are candidates to be placed in analogical correspondence. We suggest a process of re-representation during comparison by which semantic content can be decomposed, integrated or abstracted to allow for the alignment of underlying commonalities between base and target (see Gentner & Medina, in press).

There has not been a direct experimental test of how these processes occur in real time. The present investigation uses a methodological paradigm in which participants make on-line judgments about the analogical relatedness of pairs of structured stimulus items that vary in their similarity relationships. We report accuracy and RT data in the evaluation of analogies that reveal systematic differences depending on the kind and degree of similarity between items being compared. Implications of these data for the underlying process of comparison are considered.

ACKNOWLEDGMENTS:

This research was supported in part by National Science Foundation Grant BNS-9096259, Office of Naval Research Grant N00014-89-J-1272, and an NIH-NRSA postdoctoral training grant to the first author. We thank the Similarity & Analogy group at Northwestern University.

REFERENCES:

Falkenhainer, B., Forbus, K.D. & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. Artificial Intelligence, 41(1), 1-63.

Gentner, D. & Markman, A.B. (1993). Analogy ... Watershed or Waterloo? Structural alignment and the development of connectionist models of cognition. In S.J. Hanson, J.D. Cowan & C.L. Giles (Eds.), Advances in Neural Information Processing Systems, 5 (pp. 855-862). San Mateo, CA: Kaufmann.

Gentner, D. & Medina, J. (in press). Similarity and the development of rules. Cognition.

Hummel, J.E. & Holyoak, K.J. (1997). Distributed representations of structure: A theory of analogical access and mapping. Psychological Review, 104(3), 427-466.

AN EXPERIMENTAL PARADIGM TO STUDY SPONTANEOUS ANALOGIES INVOLVED IN PROBLEM SOLVING SITUATIONS

Emmanuel Sander
University of Paris 8
Department of Psychology
93526 Saint Denis Cedex 02, France

Evelyne Clement
University of Rouen
Department of Psychology
76821 Mt St Aignan Cedex, France

Experimental studies on analogy making mainly rely on a 'source-target' paradigm in which a source situation is taught to the participants before testing their behavior within the target situation. It enables to control the knowledge of the subjects concerning the source; and also to manipulate the source in order to study the influence of the manipulated features. This paradigm can also be directly transposable in teaching situations in which the source can be taught in order to help the subject understand the target. This paradigm also reveals some limits. Firstly, it is difficult to control to which extent previous knowledge intervenes in the process of building a representation of the source and of the target. Some interpretative effects have been demonstrated in those situations (Bassok, Wu, & Olseth, 1995). Secondly, this paradigm is not suitable for studying the whole range of the analogies. In ecological situations, spontaneous analogies usually rely on familiar sources which can hardly be taught within an experimental session. For instance, children take their knowledge about human beings as a source in many situations (Inagaki & Hatano, 1991).

Another paradigm can be used to study spontaneous analogies in which any knowledge in long term memory may be a potential source: no source is given to the participant and his/her behavior is compared with the one predicted through an hypothesized source. This paradigm allows to predict and explain the difficulties met by participants in a wider range of situations than within the classical paradigm.

We present two experiments in which problem solving situations are analyzed as relying on analogies with familiar sources.

In the first experiment, children who started to study column subtractions without borrowing are asked to solve column subtraction with borrowing. Their mistakes were predicted through the reference to two main familiar sources: subtracting is like taking a part from a whole, and subtracting is like covering a distance. A model was built on the basis of the use of those analogies, and the result of the simulation was compared to the pattern of responses. We are able to simulate 83% of the responses.

In a second experiment, adults are asked to solve isomorphs of the Tower of Hanoi in which they have to move or to change the size of objects. Difficulties are predicted through the use of two sources depending on the isomorph: knowledge about taking a lift, and knowledge about biological growth. We show that the difficulties result from the use of these familiar sources. Their use entails additional constraints which lead to building inadequate problem-space.

The results support the idea that analogies allow the learners to attribute to the new situations the properties of well known situations. The interest of this paradigm is that it allows to point out the nature and the functions of the familiar knowledge implied in analogy mechanism.

REFERENCES

Bassok, M, Wu, L.L., & Olseth, K.L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory and Cognition, 23*, 354-367.

Inagaki, K., & Hatano, G. (1991). Constrained person analogy in young children's biological knowledge. *Cognitive Development, 6*, 219-231.

LEARNING FROM EXAMPLES:
CASE-BASED REASONING IN CHESS FOR NOVICES

Andre Didierjean
Laboratoire Cognition & Communication
Paris V - CNRS
46, rue Saint Jacques
75005 Paris
Email: Andre.Didierjean@parisV.sorbonne.fr

Evelyne Cauzinille-Marmëche
CREPCO
Centre de Recherche en Psychologie Cognitive
Universitè de Provence - CNRS
29 avenue Robert Schuman
F-13621 Aix en Provence cedex 1

One fundamental question in cognitive psychology is whether knowledge constructed during the analysis of examples is stored in an abstract form or whether it is kept in its full form. A related question concerns the conditions under which knowledge can be used in a problem solving situation. Can an example be understood and reused to solve a new problem without resorting to an abstract representation? We present two studies, with novices in the game of chess, investigating the existence of a process of reasoning by analogy that does not require the mediation of an abstract knowledge structure. In the first experiment, subjects analyse chess problem examples and then solve similar problems. The results showed that during transfer, subjects use knowledge that has a very low degree of abstraction: they only succeed on problems similar to the examples when they are perceptually close (in particular, they failed when we changed, symmetricly, the chess pieces position on the chessboard).

Experiment 2 investigates the role of failure in analogical transfer. From the results it seems that attempting to solve the source problem, and encounter failures, is a determinant in case-based reasoning.

DEVELOPMENT OF NUMERICAL EQUIVALENCE JUDGMENTS: APPEARANCES COUNT

Kelly S. Mix
Department of Psychology
Indiana University
Bloomington, IN 47405, USA
kmix@indiana.edu

ABSTRACT

The present study investigated whether preschool children recognize numerical equivalence between sets of objects that vary in similarity. The results indicate that the ability to recognize numerical equivalence for varying object sets emerges gradually during the preschool period. Verbal counting ability is linked to success on some but not all comparisons.

BACKGROUND

On the face of it, the task of judging numerical equivalence seems much like judging similarity along any other dimension–entities are compared and a common attribute or relation is identified. Therefore, one might expect children's numerical equivalence judgments to develop like similarity judgments in other domains. For example, the effects of surface similarity on children's comparisons are well-documented in a variety of non-numerical tasks (Gentner & Toupin, 1985; Holyoak, Junn, & Billman, 1984; Kotovsky & Gentner, 1996; Rattermann, Gentner, & DeLoache, 1989). Thus, children may have difficulty recognizing number as the relevant relation when the sets being compared are otherwise very different. In addition, children's responses in numerical equivalence tasks may shift from an emphasis on surface similarity to an emphasis on relational similarity over development–i.e., the relational shift described in other domains (Gentner, 1988, Gentner & Rattermann, 1991). Finally, knowledge of the count words might improve numerical equivalence judgments just as the act of naming has helped focus children's attention on category-relevant dimensions in other domains (Gentner & Rattermann, 1991; Smith, 1993).

However, current views of number development paint a different picture. Reports of numerical abstraction in infants, as well as other early numerical competencies, have led to the proposition that numerical development is guided by a set of innate domain-specific principles (Gallistel & Gelman, 1992; Gelman, 1991). These principles are supposed to provide a structure that supports and promotes numerical development. If so, then development of numerical equivalence judgments might be immune to the difficulties children encounter judging other types of similarity.

METHOD

The basic procedure involved a triad matching task in which preschool children matched a target set with 2, 3, or 4 items to one of two choice cards that showed an equivalent number of dots. The critical manipulation was that the contents of the target sets varied across conditions. In one condition, the target sets were nearly identical to the sets on the choice cards (dots-to-dots). In a second condition, the target sets were homogeneous groups of objects that were different from the sets on the choice cards (shells-to-dots). In the third condition, the target sets were heterogeneous sets of objects that also differed from the sets on the choice cards (random objects-to-dots). In addition to these matching tasks, children also were given several counting tasks to assess their knowledge of the conventional count words.

RESULTS

There was a clear difference in performance depending on which comparison children were making. First, the conditions with less surface similarity were significantly more difficult than the literal dots-to-dots condition across age (Shells-to-dots vs. dots-to dots: F(1,28) = 8.71, $p < .01$; Random objects-to-dots vs. dots-to dots: F(1,42) = 28.74, $p < .0001$). This is consistent with work in other domains showing that surface similarity affects transfer in young children.

Second, there was evidence of a relational shift. Children performed above chance on the disks-to-dots comparison at a younger age than children performed above chance on shells-to-dots comparison. Furthermore, children performed above chance on the shells-to-dots comparison at a younger age than children performed above chance on random objects-to-dots. Thus, over development, children gradually extended their equivalence judgments from comparisons with high surface similarity to comparisons with only relational similarity.

Third, conventional counting ability appeared to improve performance. Children who were competent counters performed all three matching tasks above chance. However, children who were not competent counters performed at chance on the shells-to-dots and random objects-to-dots comparisons. Thus, knowing the verbal labels for small sets may aid in transfer for less literal numerical comparisons.

CONCLUSIONS

The present results indicate that numerical equivalence judgments develop much like other comparisons–inasmuch as surface similarity and labeling affect performance. In contrast, the present findings are inconsistent with the view that development of number concepts is privileged by virtue of innate, domain specific knowledge structures.

THE EFFECTS OF A TRAINING PROGRAM ON THE ANALOGICAL REASONING ABILITIES OF AN ELEMENTARY SCHOOL-AGED SAMPLE

Doris Johnson
Department of Psychology and Counseling
University of the District of Columbia, Washington, D.C.

Albert Roberts
Department of Psychology Howard University

Analogical reasoning, an important cognitive skill, invoved perceiving similar relationships in dissimilar domains. Early theorists believed that true analogical reasoning capability was achieved around adolescence and that young children were incapable of engaging in analogical reasoningand transfer. Analogical transfer involves ignorning nonanalogous information, extracting relevant analogous information from one particular domain, and using it to answer questions or solve problems in a different domain.

However, much recent research has demonstrated not only early analogical reasoning, but early analogical transfer abilities in children. Much new research has focused on children three, four, and five years old. However, few studies occur in the regular classrom or seek to illuminate the capabilties of elementary school-aged children. This study sought to addres these issues.

Seven 4th-grade classes, four expeirmental and three ocntrol, participated in a group intervention designed to train students in analogical reasoning and transfer. The training was undergirded by principles embraced by the knowledge-based view of analogical reasoning. This perspective holds that if children are familiar with the objects in the analogy and understand the relations between the items in the analogy, they will have no difficulty engaging in analogial solution and transfer.

The intervention consisted of six sessions: pretest, metaphorical story presentation, three training sessions, and posttest. During the story presentation, studetns read a metaphorical story that served a a tool in analogy solution and transfer. The two A groups were pretested and trained on analogies from the domains of relations (such as male/female, singleton/group, part/whole and sequence), mathematics and metaphors. The A groups were posttested on analogies from the domains of word forms (such as antonyms, synonums, palindromes and homonyms), story problem solving, and spatial relations analogies. The B groups' presentations were reversed.

Analyses revealed significant training effects for one A group and both B groups. Singificant transfer effects were demonstrated for both A groups and one B group. There were no significant gender related differences either group in the posttest domains. The training was an effective vehicle for teaching children both analogical solution and analogical transfer. Further research should be done to refine the training program, with a goal of implementation in elementary schools.

ANALOGICAL GENERALISM: AN ANALOGICAL PERSPECTIVE ON THE EVOLUTION OF LANGUAGE

Jamie Carnie
96 Carlingcott, nr Bath, BA2 8AW, United
Kingdom email: adsjrc@bath.ac.uk

ABSTRACT

This paper considers from a philosophical perspective the idea that analogy has been the principal underlying mechanism in the evolutionary development of language.

The view that language is fundamentally analogical in nature is increasingly being considered by philosophers. Evidence for this view is briefly set out, in the form of the widespread systems of analogy and metaphor recently documented by Lakoff and Johnson, and of vocabulary itself, the bulk of which shows signs of having been formed by analogy-like processes of construction. The evidence is substantial enough to prompt the hypothesis on which the paper centres, that such analogical construction has been the dominant evolutionary process in language.

The paper proceeds to examine the philosophical implications of this idea, and in the course of so doing develops a theory of language evolution called Analogical Generalism which takes the idea as one of its central concepts. In considering Analogical Generalism the paper does not concern itself with individual historical languages, but rather the overall trends of language evolution which the theory implies and which, if the theory is correct, must have been instantiated in the actual development of all historical languages.

A concept of 'articulation' is introduced. This is the characteristic of words which makes some display more structure in expressing meaning than others. It is argued that some words are 'articulatively general', having little or no expressive structure, while others are 'articulatively complex'. This concept is related to the complexity of the unconscious linguistic knowledge users bring to understanding the sense of words.

It is thenb argued that analogical construction of new vocabulary can only give rise to words of greater articulative complexity than their source terms. This means that a language evolution dominated by this process must have developed broadly from articulatively general terms towards more precise, articulatively complex ones.

An evolutionary trend towards increasing expressive complexity over time implies that language must have had its origins in articulatively highly general terms. This concept introduces the 'Generalist' component of the theory developed in the paper. It is argued that there exist even in modern languages certain words of absolutely minimal articulative structure. These 'primal words' can typically be substituted for by gestures, and as such may represent a missing link between animal communication and modern human language. It is argued that examples of mammalian communication such as the barking of dogs can plausibly be thought of as expressing meaning at he same minimal level of articulation as human primal terms - indeed that in certain cases the meaning expressed may itself be identical to that expressed by human primal words.

Other aspects of the Generalist position about language origins are explored, and contrasted with the more conventional picture which sees articulation as an invariable constant in language. In various ways it is shown that this new approach represents a superior position to the rather naive Articulative Atomism of the latter. This is most particularly so in the fact that it is not committed to any radical discontinuity in the early development of mean-

ingful language. The Generalist view makes it possible to understand the evolution of the earliest words as end products of a continuous and progressive development of the primal language forms of higher primates and early hominids.

The Generalist origins that are implied by the trends which would be imposed on language development by a dominant process of analogical construction solve, then, some of the more intransigent problems concerning the origins of language. It is concluded that the hypothesis of the dominance of analogical construction, together with a Generalist account of language origins, is from a philosophical perspective sound. It is therefore proposed that the outline of a coherent account of language evolution has become evident in the theory of Analogical Generalism.

METAPHORS IN MIND AND DISCOURSE: PATTERNS OF EUPHEMIZATION

Elena Andonova
Department of Cognitive Science
New Bulgarian University
21, Montevideo Str.
Sofia 1635, Bulgaria
elena@cogs.nbu.acad.bg
elan@biscom.net

Research within the paradigms of conceptual metaphors and discourse analysis is brought together in this study of the patterns and strategies of euphemization of 'death' as a taboo topic. The phenomenon of euphemization traditionally considered from an isolated lexico-semantic point of view is explored here within a combined model of metaphoric patterns and discursive strategies. Conceptualization of 'death' is carried out along a number of dimensions such as: individual vs. universal experience, controlled vs. uncontrolled, irreversible vs. reversible, gradual vs. sudden (expected vs. unexpected), event vs. state, etc. It is based on a range of conceptual metaphors–ontological, structural, orientational. The choice of metaphorical pattern highlighting certain orientations within the various dimensions serves euphemistic purposes. Thus, euphemistic is the preferred use of one underlying conceptual metaphor instead of another in the construal of the concept of death (e.g., Death-as-Journey vs. Death-as-Struggle). Discourse structure is examined in texts employing a set of strategies which exploit certain aspects of conceptual structure as identified above for purposes related to the psychological motivation of the usage of euphemization (general models of human coping behaviour) as well as communicative goals which reflect situational characteristics, e.g., text genre. Different discourse-framing devices are used. Thus, the study reveals the existence of a systematic relationship between the patterns of selective highlighting of conceptual structure and discourse constructive strategies which constitute euphemization as a psychologically and communicatively motivated phenomenon.